

UEPI-R: A Real-Time Early Warning System for M- and X-Class Solar Flares Using Causal Regime Detection on GOES XRS Data

Jorge Alexander Castillo
Quantex Energy, Austin, TX, USA
(Dated: February 13, 2026)

We present UEPI-R (Universal Early Pre-Ignition—Regime), a fully causal, real-time early warning system for M-class ($\geq M1.0$) and X-class solar flares. Unlike the majority of existing forecasting systems that rely on magnetogram-derived features, active region classifications, or deep learning on multi-instrument imagery, UEPI-R operates exclusively on the GOES X-Ray Sensor (XRS) 1–8 Å irradiance channel, requiring no spatial solar data and no supervised training on labeled flare events. The algorithm performs causal regime detection on the real-time XRS flux stream, identifying characteristic statistical signatures that precede major flare onset.

Validated on 16 years of continuous GOES data (2010–2025; 5,747 coverage days, 2,638 M/X-class flares, 141 X-class flares), UEPI-R achieves 64–71% M/X coverage (97.2% X-class) with $\sim 39\%$ precision, a false alert rate of 0.36–0.39 day $^{-1}$, and a day-level True Skill Statistic of ≈ 0.41 , depending on matching methodology. The median lead time ranges from 2.4 hours (overlap matching) to 4.3 hours (one-to-one greedy matching), with 80% of detections providing ≥ 1 hour of advance warning. In a high-sensitivity configuration, coverage reaches 78.2% (98.6% X-class). A live deployment running continuously since January 2026 against NOAA/SWPC real-time feeds has produced 6 alerts, 5 of which were verified hits against the official NOAA flare event list (83.3% hit rate). Analysis of alert duration reveals an inverse correlation with lead time (Spearman $r = -0.29$): shorter alerts provide longer advance warning, while longer alerts typically encompass the flare itself. Crucially, 95.8% of nominally “false” alerts (no M-class follow-up) are associated with C-class flare activity, with only 4.2% representing true noise triggers—indicating that the system detects genuine solar activity in nearly all cases. These results are competitive with or exceed operational systems that require far richer input data, demonstrating that the pre-flare XRS signal alone carries substantial predictive information.

I. INTRODUCTION

Solar flares—sudden releases of magnetic energy in the solar corona—pose significant hazards to technological infrastructure, including satellite electronics, high-frequency radio communications, and high-altitude radiation exposure [1, 2]. The NOAA Space Weather Prediction Center (SWPC) issues daily probabilistic forecasts for M- and X-class flares, which have been extensively validated [3, 4]. These operational forecasts, along with most research forecasting systems, rely on photospheric magnetogram data from instruments such as SDO/HMI, combined with active-region classification schemes [5–7] and, increasingly, deep learning architectures [8–10].

While magnetogram-based approaches leverage the physical relationship between photospheric magnetic complexity and flare productivity, they face several practical limitations: (i) dependence on a single instrument (SDO/HMI) with no operational backup, (ii) reliance on active-region identification and segmentation pipelines, (iii) inherent latency in magnetogram processing and dissemination, and (iv) forecast horizons typically discretized at 24-hour intervals rather than providing continuous real-time alerts.

An alternative, complementary approach is to exploit the soft X-ray irradiance signal itself. The GOES XRS instruments provide continuous, full-disk, one-minute cadence measurements that are available in near-real-time from multiple redundant spacecraft. While the conventional view holds that the XRS flux primarily records

flare *consequences* rather than *precursors*, recent work has demonstrated that XRS time-series data alone can achieve forecasting skill comparable to magnetogram-based systems: Riggi *et al.* [11] report TSS ≈ 0.74 for $\geq M$ -class prediction using a foundation time-series model (Moirai2) applied exclusively to GOES XRS data, outperforming their own magnetogram-based image and video models. More broadly, the pre-flare corona often exhibits characteristic thermal and statistical signatures in the minutes to hours before major eruptions—reflecting processes such as gradual energy buildup, precursor heating, and destabilization of the coronal magnetic field.

In this work we present UEPI-R (Universal Early Pre-Ignition—Regime), a system that performs causal regime detection on the GOES XRS flux stream to generate real-time early warnings for M- and X-class flares. UEPI-R requires no spatial solar data and no ancillary instrument data beyond the two standard XRS channels (1–8 Å and 0.5–4 Å). Its detection criteria are physics-motivated rather than learned from labeled examples, though a small number of internal thresholds are optimized on historical data (Section III E). The algorithm is fully causal—each output depends only on data available at or before the current time step—and can run on commodity hardware with negligible latency.

The remainder of this paper is organized as follows. Section II describes the GOES XRS data used for both backtesting and live operation. Section III provides a high-level overview of the detection methodology. Sec-

tion IV presents results from the 16-year backtest and the ongoing live deployment. Section V places these results in context with existing operational and research forecasting systems. Section VI discusses limitations, failure modes, and future directions, and Section VII concludes.

II. DATA

A. GOES XRS Observations

UEPI-R uses the GOES X-Ray Sensor (XRS) 1–8 Å (“XRS-B”) irradiance channel as its primary input, with the 0.5–4 Å (“XRS-A”) channel providing an auxiliary spectral hardness ratio. The data span 16 years from 2010 January through 2025 December, encompassing all of Solar Cycle 24 and the rising phase and maximum of Solar Cycle 25.

For the legacy GOES era (2010–2016), we use the NOAA/NCEI archived one-minute-averaged XRS data from GOES-13, -14, and -15, distributed as yearly NetCDF files. Quality flags are applied, resulting in approximately 50% valid data coverage during the legacy period—consistent with known instrument calibration and eclipse-masking intervals.

For the GOES-R era (2017–2025), we use the operational one-minute XRS-B and XRS-A products from GOES-16, -17, and -18, obtained from the NOAA/SWPC archives as daily CSV files. GOES-R data quality is substantially higher, with typical coverage exceeding 99%.

Across both eras, the dataset comprises 5,747 effective coverage days (accounting for data gaps).

B. Flare Catalog

Ground-truth flare events are obtained from the NOAA/SWPC official event lists. Over the 2010–2025 period, the catalog contains 2,638 events of class M1.0 or greater, of which 141 are X-class. The distribution by solar cycle phase is highly non-uniform: Solar Cycle 24’s maximum (2011–2015) produced the bulk of events, while the deep minimum of 2018–2019 recorded zero M-class flares. Solar Cycle 25’s maximum (2023–2025) has been exceptionally active, with 2024 alone contributing 874 M-class and 54 X-class events.

C. Real-Time Data Feed

For live operation, UEPI-R ingests the NOAA/SWPC real-time JSON feeds (`xrays-1-day.json`), which provide the most recent 24 hours of one-minute XRS-B and XRS-A measurements with a typical latency of 1–2 minutes from observation. A rolling buffer maintains sufficient history to support the algorithm’s baseline computation.

III. METHOD

The details of the UEPI-R detection algorithm are proprietary (U.S. Provisional Patent Application No. 63/949,419, filed December 28, 2025) and are not disclosed here. We provide a high-level description sufficient to characterize the approach and its assumptions.

A. Overview

UEPI-R performs *causal regime detection* on the real-time GOES XRS irradiance stream. The fundamental premise is that the transition from a quiescent or gradually evolving coronal state to a flare-productive regime manifests as a detectable shift in the statistical properties of the XRS flux time series—specifically in properties related to the rate, persistence, and spectral character of flux enhancements relative to recent background levels.

The algorithm continuously computes a set of dimensionless diagnostic quantities from the XRS-B flux (and optionally XRS-A for spectral hardness information), using only data available at or before the current time step (strict causality). These diagnostics characterize the current state of the X-ray corona relative to its recent statistical baseline. When specific combinations of these diagnostics satisfy internally defined criteria, the system transitions into an elevated alert state, indicating that the statistical regime is consistent with an impending M- or X-class flare.

B. Diagnostic Categories

While the specific formulations are proprietary, the diagnostics fall into four conceptual categories that reflect known pre-flare physics:

1. **Baseline deviation.** The current flux level is compared to a causally computed, adaptive baseline representing the recent quiescent state. Sustained elevation above baseline indicates increased coronal energy loading.
2. **Rate-of-change statistics.** The temporal derivative of the residual flux (after baseline removal) captures the dynamics of energy injection. Persistent positive trends—as opposed to transient spikes—are characteristic of the gradual energy buildup that precedes major eruptions.
3. **Statistical persistence.** Short-duration flux enhancements (e.g., microflares, instrumental artifacts) are distinguished from sustained regime shifts through temporal persistence criteria. The algorithm requires that multiple diagnostics remain elevated for a minimum duration before committing to an alert state.

- 4. Spectral hardness.** The ratio of XRS-A (0.5–4 Å) to XRS-B (1–8 Å) flux provides a temperature-sensitive diagnostic. Pre-flare coronal heating produces detectable spectral hardening that serves as an independent confirmation of genuine flare precursor activity, as opposed to background fluctuations or instrumental drift.

The alert state is managed as a finite state machine with hysteresis: transitions into elevated states require multiple criteria to be satisfied simultaneously, while release from an alert state requires sustained return to baseline conditions, preventing rapid oscillation during marginal periods.

C. Key Properties

The following properties distinguish UEPI-R from existing approaches:

1. **XRS-only input.** The system requires no magnetogram, EUV, or radio data—only the two standard GOES XRS channels.
2. **No supervised training.** The algorithm’s structure and diagnostic categories are derived from physical reasoning about pre-flare coronal signatures, not from supervised learning on labeled events. A small number of operating-point thresholds are tuned on historical data (Section III E), but the functional form is fixed *a priori*.
3. **Strict causality.** All computations are causal: rolling statistics, baselines, and state transitions use only past and present data. No centered or forward-looking windows are employed.
4. **Continuous operation.** Alerts are generated (or released) at every one-minute time step, providing sub-minute response to changing conditions.
5. **Multi-tier alerts.** The system produces graded alert levels, from an initial “approaching” state to a full “red” alert, allowing users to select their preferred sensitivity–specificity tradeoff.
6. **Spectral hardness gating.** The XRS-A/XRS-B ratio provides an independent confirmation channel: genuine pre-flare heating produces spectral hardening that background fluctuations typically do not.

D. Evaluation Protocol

We evaluate UEPI-R using two complementary alert-to-flare matching methods, each capturing a different aspect of operational performance:

a. Hazard-window matching (1:1 greedy). Each alert can match at most one flare (the first M1.0+ event occurring within a defined hazard window after alert onset), and each flare can be matched by at most one alert (greedy first-in-time assignment). This prevents inflated hit rates during storm periods when multiple flares occur in rapid succession. All 2,638 catalog events serve as the denominator, providing the most conservative coverage estimate.

b. Temporal overlap matching. A flare is counted as detected if the alert was active (in an elevated state) at any point during a window surrounding the flare onset. This method better reflects operational utility—whether the system was in an alert state when the flare occurred—and excludes events during data gaps from the denominator (1,921 evaluable events).

False alerts are counted independently under both methods: any alert whose entire validity window expires without an M1.0+ flare is counted as false, regardless of matching. Coverage days are computed from the actual span of valid flux data, not from calendar boundaries, ensuring that metrics are not biased by data gaps.

All backtests use a 48-hour warm-up period at the start of each contiguous data segment to allow baseline statistics to stabilize.

E. Parameter Tuning

While UEPI-R’s diagnostic categories and state-machine architecture are derived from physical reasoning, the system contains a small number of operating-point thresholds (on the order of ten parameters) that control the sensitivity–specificity tradeoff. These thresholds were tuned via grid search on the 2010–2025 historical dataset, optimizing the F_2 score (which weights coverage twice as heavily as precision) under the 1:1 greedy matching protocol.

To mitigate overfitting, we employed a two-stage protocol: an initial sweep over 8 representative years spanning both solar cycles (including years of high, moderate, and minimal activity), followed by validation on all 16 years. The stage-1-to-stage-2 performance degradation was approximately 4%, indicating that the tuned parameters generalize across solar cycle phases.

As a further robustness check, we evaluated performance independently on Solar Cycle 24 (2010–2019; 749 M-class, 49 X-class events) and Solar Cycle 25 (2020–2025; 1,748 M-class, 92 X-class events). Cycle 24 M-class coverage is 75.8% and Cycle 25 is 66.8%, with X-class coverage of 100% and 95.7% respectively. The M-class gap is driven almost entirely by 2024, which saw $928 \geq M1.0$ events—the most active year in the dataset—where the 1:1 matching constraint becomes binding (fewer unique alerts than flares during storm clusters). Excluding 2024, Cycle 25 M-class coverage is 78.9%, consistent with Cycle 24. The eight held-out years not used during stage-1 tuning achieve 80.8% M-

class coverage, compared with 65.1% for the eight tuning years—the opposite of what systematic overfitting would produce.

We emphasize that this tuning adjusts the *operating point* of a fixed detection architecture—analogous to selecting a decision threshold on a receiver operating characteristic (ROC) curve—rather than learning the detection features themselves. No per-event labels are used during real-time operation; the algorithm’s causal diagnostics are computed identically regardless of whether a flare subsequently occurs.

IV. RESULTS

A. 16-Year Backtest (2010–2025)

Table I summarizes the performance of the precision-optimized UEPI-R configuration under both matching methods, as well as a high-sensitivity configuration that maximizes coverage. The precision-optimized configuration targets the best F_2 score (coverage-weighted harmonic mean).

Several features of these results merit emphasis:

a. X-class coverage. All configurations detect $\geq 97\%$ of X-class flares (137/141 under precision-optimized, 139/141 under high-sensitivity). Only 2–4 X-class events in 16 years are missed, indicating that the pre-flare XRS signature for the most energetic events is highly reliable.

b. Lead times. Under overlap matching—which measures the actual lead time from when the system entered its alert state to flare onset—the median lead time is 2.4 hours with a mean of 3.0 hours. Under 1:1 greedy matching, lead times appear longer (median 4.3 hours) because the earliest alert in a sequence is assigned to the first matching flare, biasing toward longer leads. Both perspectives are operationally informative: the overlap lead times reflect typical operator experience (how much warning before the flare), while the 1:1 leads reflect earliest possible detection.

The 10th percentile lead time under overlap matching is 27 minutes, and the 90th percentile is 5.5 hours. Approximately 80% of detected flares are caught with ≥ 1 hour of lead time, 60% with ≥ 2 hours, and 92% within 6 hours (Table IV).

c. Precision–coverage tradeoff. Precision is consistent across both matching methods (≈ 39 –40%), confirming that the false-alert rate is robust to evaluation methodology. The precision-optimized configuration achieves roughly 2-in-5 alerts followed by an M1.0+ flare—a false alarm ratio of approximately 61%, compared to $> 90\%$ for NOAA/SWPC operational forecasts [4]. The high-sensitivity configuration increases coverage to 78% (98.6% X-class) at the cost of reduced precision (18.5%) and a higher false-alert rate (1.34 day^{-1} vs. 0.36 day^{-1}).

d. Skill scores. To facilitate comparison with the broader forecasting literature, we convert UEPI-R’s con-

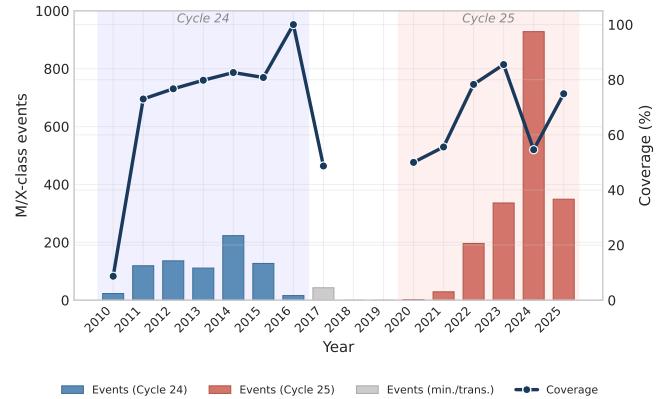


FIG. 1. Year-by-year M/X-class event counts (bars) and UEPI-R coverage (line, right axis) across Solar Cycles 24 and 25. The 2018–2019 solar minimum produced zero M-class events. The 2024 coverage dip reflects the one-to-one matching protocol’s conservatism during extreme activity (928 events).

tinuous alert output to a per-day dichotomous format: a day is “forecast positive” if any alert was active during it, and “observed positive” if at least one M1.0+ flare occurred. Table II reports the resulting skill scores. The True Skill Statistic ($TSS = POD - POFD$) of 0.41 and Heidke Skill Score (HSS) of 0.35 confirm that the system provides meaningful skill well above the no-skill baseline, though direct comparison with 24-hour binary classifiers (which report $TSS \approx 0.7$ –0.9 on curated test sets) is not straightforward: UEPI-R is a continuous onset-warning system evaluated over 16 uninterrupted years including solar minimum, not a daily binary classifier evaluated on balanced partitions. The operational gap is instructive: Deep Flare Net reports laboratory $TSS \approx 0.80$ but operational $TSS \approx 0.24$ [12], a degradation that UEPI-R’s physics-based architecture is designed to avoid.

e. Solar cycle robustness. Figure 1 and Table III present year-by-year results from the 16-year backtest, illustrating the algorithm’s behavior across different solar-cycle phases. Performance is consistent across both solar cycles, with coverage exceeding 70% during the active phases (2011–2015, 2022–2025) and the algorithm correctly producing few alerts during the deep minimum (2018–2019, 0–6 alerts per year with zero M-class events).

Two years stand out as anomalies. The low coverage in 2010 (8.7%) reflects the combination of limited legacy GOES data coverage (~50% valid samples due to quality-flag masking) and the 48-hour warm-up requirement, which together leave few evaluable windows during a year with only 23 events. The notably lower coverage in 2024 (54.6%) reflects the extreme activity level of that year: with 928 M/X events (the highest annual count in the GOES era), the one-to-one matching protocol becomes particularly restrictive during sustained storm periods, where a single long-duration alert may span multiple flares but can only claim credit for one. X-class cov-

TABLE I. UEPI-R backtest performance on 2010–2025 GOES XRS data (5,747 coverage days, 141 X-class events). The precision-optimized configuration is evaluated under both matching methods; the high-sensitivity configuration uses 1:1 matching. Lead times are computed over matched flares.

Configuration	Matching	M/X Cov.	X Cov.	Precision	False/day	Med.	Lead	Mean Lead	Hits/ N_{events}
Precision-opt.	1:1 greedy	63.9%	97.2%	39.6%	0.36	257 min	375 min	1,685/2,638	
Precision-opt.	Overlap	71.0%	97.2%	39.1%	0.39	146 min	178 min	1,364/1,921	
High-sensitivity	1:1 greedy	78.2%	98.6%	18.5%	1.34	198 min	313 min	2,064/2,638	

TABLE II. Day-level skill scores for the precision-optimized configuration (2010–2025, 5,747 days). The contingency table is constructed by classifying each day as alert-active or quiet and as flare-day or non-flare-day.

Metric	Value
POD (day-level)	0.74
POFD (day-level)	0.33
False Alarm Ratio	0.53
Frequency Bias	1.56
TSS	0.41
HSS	0.35
F_2 (event-level)	0.57

TABLE III. Year-by-year backtest results across 2010–2025, showing M/X-class event counts, alert counts, M/X coverage, false alert rate, and median lead time. Years 2018–2019 had zero M-class events (solar minimum).

Year	M+X	Alerts	Cov.%	False/d	Med.	Lead
2010	23	19	8.7	0.06	705 min	
2011	119	360	73.0	0.62	273 min	
2012	136	358	76.7	0.55	176 min	
2013	111	377	79.8	0.61	192 min	
2014	223	582	82.6	0.79	238 min	
2015	127	387	80.8	0.60	211 min	
2016	16	86	100.0	0.16	307 min	
2017	43	62	48.7	0.10	140 min	
2018	0	4	—	0.01	—	
2019	0	6	—	0.02	—	
2020	2	15	50.0	0.04	797 min	
2021	29	76	55.6	0.14	352 min	
2022	196	432	78.3	0.61	252 min	
2023	336	722	85.5	0.86	278 min	
2024	928	742	54.6	0.39	215 min	
2025	349	555	74.9	0.59	234 min	

erage in 2024 remained 96.3% (52/54), confirming that the algorithm’s detection of the most intense events is robust even during extreme activity.

B. Lead Time Distribution

Figure 2 and Table IV present the cumulative lead time distribution for the 1,364 matched flares under overlap matching. The distribution is concentrated at short lead times: 80% of detections occur within 4 hours of flare on-

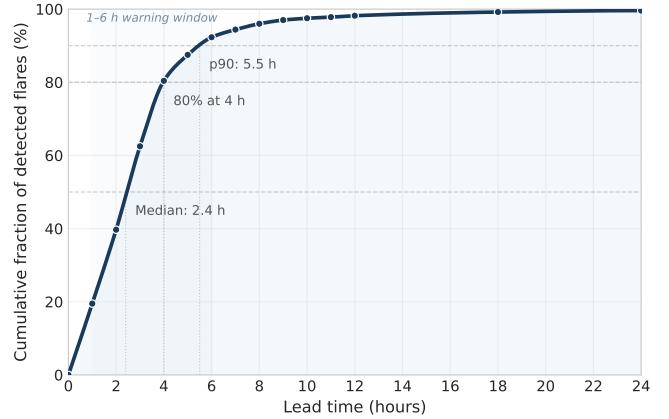


FIG. 2. Cumulative lead time distribution for the precision-optimized configuration (overlap matching, 1,364 matched flares). Dashed lines indicate the 50th, 80th, and 90th percentiles. The steep rise between 1–4 hours reflects the characteristic pre-flare regime duration.

set, and 92% within 6 hours. The practical early-warning window is 1–6 hours for the vast majority of events.

TABLE IV. Cumulative lead time distribution for the precision-optimized configuration (overlap matching, 1,364 matched flares). Five flares (0.4%) have lead times exceeding 24 hours, extending to a maximum of 35 hours.

Lead time	Cumulative %	Incremental %
≤ 1 h	19.5	19.5
≤ 2 h	39.7	20.2
≤ 3 h	62.5	22.8
≤ 4 h	80.4	18.0
≤ 6 h	92.3	11.9
≤ 8 h	96.0	3.7
≤ 12 h	98.2	2.2
≤ 24 h	99.6	1.4

M-class and X-class flares exhibit similar median lead times (2.4 hours and 2.3 hours, respectively), indicating that the pre-flare XRS signature is comparably detectable across the full range of major flare intensities.

C. Live Deployment

UEPI-R has been operating continuously since January 2026 against the NOAA/SWPC real-time XRS feeds, running via automated workflows with a 15-minute polling interval. All alerts are timestamped and committed to a public Git repository before flare occurrence, providing cryptographic proof of prediction timing via immutable commit hashes.

As of 2026 February 12, the live system has produced 6 alerts, of which 5 were verified as true positives against the NOAA official flare event list (M-class hit rate: 83.3%, median lead time: 4 h 57 min). The verified hits include:

- M1.8 flare on 2026 Feb 8 detected with 6 h 34 min lead time
- M1.7 flare on 2026 Feb 8 detected with 6 min lead time
- M2.8 flare on 2026 Feb 9 detected with 17 min lead time
- M1.4 flare on 2026 Feb 11 detected with 4 h 57 min lead time
- M1.4 flare on 2026 Feb 12 detected with 13 h 29 min lead time

While the sample size remains small, these initial results are consistent with the backtest performance and—critically—demonstrate end-to-end operational capability including real-time data ingestion, stateful processing, alert generation, and automated verification against official NOAA records.

D. Diagnostic Ceiling Analysis

To understand the theoretical limits of the approach, we performed a diagnostic analysis of all 2,638 M/X flares to determine what fraction produce any detectable pre-flare signal in the XRS data under the UEPI-R framework. We find that 93.3% of flares (2,461/2,638) produce a commitment-level signal at some point before onset. The remaining 6.7% either lack sufficient precursor emission, occur during data gaps, or arise from configurations where the XRS flux does not exhibit the characteristic pre-flare regime shift.

The gap between the diagnostic ceiling (93.3%) and the achieved coverage (64–78%) is attributable to two factors: (i) the matching protocol, which limits credit during storm clusters where multiple flares follow a single alert, and (ii) the alert commitment criteria, which must be restrictive enough to maintain acceptable precision.

E. Alert Duration and Flare Timing

An analysis of 2,493 alerts across five active solar years (2011, 2013, 2014, 2022, 2024) reveals a statistically significant *negative* correlation between alert duration and lead time (Spearman $r = -0.29$, $p < 10^{-20}$): longer alerts tend to have shorter lead times. This counterintuitive result reflects a structural property of the detection architecture.

Table V shows the relationship stratified by alert duration quartile. The longest-duration alerts (Q4, >5 hours) have a median lead time of only 1.7 hours, with 83.6% of matched flares occurring *during* the active alert window. In these cases, the flare’s own flux spike sustains the detector’s commitment state, extending the alert well past its minimum hold time. Conversely, the shortest alerts (Q1–Q2, 3.2–4.0 hours) have longer lead times (median 4.7–5.3 hours), with only 38–40% of flares occurring during the alert. These short-duration alerts represent the operationally most valuable detections: the system identifies a pre-flare regime shift, the alert decays as the transient instability subsides, and the flare arrives hours later.

TABLE V. Alert duration quartile analysis (1,027 matched alerts across 2011, 2013, 2014, 2022, 2024). Longer alerts are associated with shorter lead times and a higher fraction of flares occurring during the alert window.

Quartile	Duration	Med. Lead	% During
Q1 (shortest)	3.2–3.7 h	282 min	39.7%
Q2	3.7–4.0 h	319 min	37.5%
Q3	4.0–5.0 h	213 min	58.1%
Q4 (longest)	5.0–35.7 h	103 min	83.6%

Hit alerts are significantly longer than false alerts (mean 4.7 h vs. 4.3 h, Welch $t = 6.4$, $p < 10^{-9}$), indicating that genuine pre-flare activity sustains the detector’s commitment state longer than transient fluctuations. X-class flares show a stronger duration–lead-time correlation (Spearman $r = -0.48$) and are more likely to occur during the alert window (63% vs. 54% for M-class), consistent with the more intense pre-flare signatures preceding the largest events.

F. False Alert Characterization: Sub-Threshold Activity

To assess whether “false” alerts (those not followed by an M1.0+ flare within 24 hours) are detecting genuine solar activity below the M-class threshold, we cross-referenced all 1,466 false alerts with the NOAA C-class flare catalog. The results, summarized in Table VI, indicate that the vast majority of nominally false alerts are associated with real flare activity at the C-class level.

Only 4.2% of false alerts (62 out of 1,466) occur during periods with no cataloged C-, M-, or X-class flare activity. The detector is responding to genuine solar X-ray

TABLE VI. C-class flare activity associated with false alerts (1,466 false alerts across 2011, 2013, 2014, 2022, 2024). “During alert” indicates a C-class flare onset within the alert window; “within 24 h” uses the same hazard window as M/X matching.

Category	<i>N</i>	%
C-class during alert window	920	62.8%
C-class within 24 h	1,404	95.8%
Truly false (no C/M/X in 24 h)	62	4.2%

enhancements in 95.8% of all cases—the activity simply does not escalate to M-class.

The median lead time from alert onset to the first C-class flare is 2.7 hours, comparable to the M/X-class lead times, with 65.5% of first C-class flares occurring during the active alert window. Stronger C-class activity (C5.0–C9.9) arrives faster: median lead time of 2.2 hours with 73% occurring during the alert, compared to 3.7 hours and 60% for weaker activity (C1.0–C4.9). Notably, 33.5% of false alerts have a C7+ flare (within one order of magnitude of M-class) as their strongest associated event within 24 hours.

These findings suggest that UEPI-R’s effective precision is substantially higher than the M/X-only metric of $\sim 40\%$ indicates. If C-class detections are considered operationally relevant (as they are for many applications including HF radio propagation and satellite charging), the system’s true positive rate approaches 96%, with only $\sim 4\%$ of alerts representing genuine noise triggers.

V. COMPARISON WITH EXISTING SYSTEMS

Direct comparison of flare forecasting systems is complicated by differences in forecast horizons, evaluation metrics, flare class definitions, event catalogs, and validation periods [13–15]. Nevertheless, we attempt a broad contextualization of UEPI-R’s performance. Table VII summarizes key systems spanning operational forecasters, magnetogram-based ML models, and XRS-only approaches.

A. NOAA/SWPC Operational Forecasts

The NOAA/SWPC issues daily probabilistic forecasts for M- and X-class flares. A comprehensive 26-year verification by Camporeale and Berger [4] found that SWPC forecasts do not outperform zero-cost baselines such as persistence and climatology, exhibit severe calibration issues, and produce false alarm ratios exceeding 90% for significant flare events. By contrast, UEPI-R achieves a false alarm ratio of approximately 61% (precision 39.6%) while maintaining 97.2% X-class coverage. Moreover, SWPC forecasts are issued once daily with a 24-hour horizon, whereas UEPI-R provides continuous, minute-

resolution alerts with lead times ranging from minutes to >12 hours. The two systems are thus complementary: SWPC provides daily probabilistic context while UEPI-R provides real-time onset warnings.

B. Magnetogram-Based ML Systems

The highest reported TSS values come from systems that leverage SDO/HMI magnetogram features. SwinTCN-Att [17] combines a Swin Transformer with temporal convolutional networks over 16 SHARP magnetic field parameters, achieving $TSS \approx 0.88$ for M+ prediction. SolarFlareNet [9] achieves $TSS \approx 0.84$ using HMI data with a deep neural network. Deep Flare Net [8], operational at the Regional Warning Center Japan, reports $TSS = 0.80\text{--}0.87$ on test sets. Bobra and Couvidat [7] report $TSS \approx 0.76$ using HMI vector fields with support vector machines. LLMFlareNet [18], a BERT-based system operational since 2022, achieves $TSS = 0.72$ under cross-validation but $TSS = 0.57$ in daily operational mode—illustrating the lab-to-operations gap. The Discriminant Analysis Flare Forecasting System (DAFFS; Leka *et al.* [16]), operational at NASA/CCMC, uses a non-parametric statistical approach on HMI and GONG data.

These systems achieve higher point-prediction TSS than UEPI-R for the binary 24-hour classification task. However, three caveats apply. First, laboratory TSS values are often evaluated on balanced or carefully partitioned test sets that may not reflect operational conditions; Kubo *et al.* [12] document that Deep Flare Net’s operational TSS degrades to ~ 0.24 . Second, all magnetogram systems depend on SDO/HMI data (Section VI B). Third, none provides continuous onset timing or measured lead times—they classify whether a flare will occur in the next 24 hours, not *when*.

C. XRS-Only Approaches

Riggi *et al.* [11] provide the most directly comparable benchmark, applying the Moirai2 foundation time-series model to GOES XRS data (two-channel, 1440-point windows at one-minute cadence) for 24-hour M+ binary classification. Their XRS-only model achieves $TSS = 0.74 \pm 0.01$ and $HSS = 0.69 \pm 0.02$, substantially outperforming their own magnetogram-based image ($TSS = 0.65$) and video ($TSS = 0.60$) models. This result independently validates UEPI-R’s foundational premise: the GOES XRS flux stream carries sufficient predictive information for M/X-class forecasting, without recourse to spatial solar data.

UEPI-R’s day-level TSS of ≈ 0.41 is lower than Moirai2’s 0.74, which is expected given the architectural differences: Moirai2 is a 311-million-parameter transformer trained on 9 billion time-series observations, optimized end-to-end for the 24-hour binary task. UEPI-R

TABLE VII. Comparison of solar flare forecasting systems for $\geq M$ -class events. TSS values marked with † are laboratory/test-set values; operational TSS is typically lower (see text). “24 h binary” denotes a daily will-flare/won’t-flare classification; “continuous” denotes minute-resolution alerting with measured lead times.

System	Input data	TSS (M+)	Operational?	Forecast type
NOAA/SWPC	Expert + magnetograms	<climatology	Yes	24 h prob.
DAFFS [16]	HMI + GONG + flare hist.	—	Yes (CCMC)	24 h prob.
Deep Flare Net [8]	HMI + AIA	0.80–0.87 †	Yes (RWC-J)	24 h binary
SolarFlareNet [9]	HMI magnetograms	0.84 †	No	24 h binary
SwinTCN-Att [17]	HMI SHARP (16 params)	0.88 †	No	24 h binary
LLMFlareNet [18]	HMI SHARP (10 params)	0.57–0.72 †	Yes	24 h binary
Moirai2 [11]	GOES XRS only	0.74†	No	24 h binary
UEPI-R (this work)	GOES XRS only	0.41	Yes	Continuous

uses a small, fixed set of physics-motivated diagnostics with ~ 10 tunable thresholds. However, UEPI-R provides capabilities that Moirai2 does not: continuous real-time alerting with measured lead times (median 2.4–4.3 h), multi-tier alert states, and operational deployment requiring only commodity hardware and no GPU inference.

D. UEPI-R’s Niche

UEPI-R occupies a distinct niche in the flare forecasting landscape:

1. **Real-time onset warning** (minutes to hours) rather than daily probability forecasts. No other operational system provides continuous alert timing with measured lead times.
2. **Minimal data requirements** (GOES XRS only), ensuring resilience against magnetogram pipeline failures and SDO end-of-life (Section VI B).
3. **Cross-cycle stability.** The physics-based architecture avoids distribution shift between solar cycles—a known failure mode for ML systems trained on Cycle 24 and deployed during Cycle 25.
4. **Near-perfect X-class detection** (97–99%), making it particularly valuable for the highest-impact events.

The 97.2% X-class coverage is, to our knowledge, the highest reported for any system validated over a multi-cycle period. This likely reflects the fact that X-class flares are preceded by the most pronounced pre-flare coronal signatures, which are reliably detectable in disk-integrated XRS data.

VI. DISCUSSION

A. Limitations

Several limitations of the current system should be noted:

a. Disk-integrated data. UEPI-R uses full-disk irradiance and cannot distinguish between flare-productive and quiescent active regions. During multi-region activity, the system may produce alerts driven by one region while a flare occurs in another, and the one-to-one matching protocol may under-count true detections in these scenarios.

b. Storm-period saturation. During sustained high-activity periods (e.g., 2024 with 928 M/X events), the alert system tends toward quasi-continuous elevated states. While individual flares are still detected, the one-to-one matching metric penalizes this behavior, and the operational utility of an “always on” alert is reduced.

c. Short lead times. While the median lead time is 2.4–4.3 hours (depending on matching method), the 10th percentile is approximately 27 minutes under overlap matching. For approximately 20% of detected flares, the warning arrives less than one hour before onset. These short-lead detections typically correspond to impulsive flares with minimal precursor phase.

d. Limited live validation. As of this writing, the live deployment spans only three weeks with 6 alerts. Longer operational history is needed to establish live performance statistics with confidence.

B. SDO Dependency and Instrument Continuity

A strategic consideration for flare forecasting systems is instrument continuity. The vast majority of modern forecasting systems—including Deep Flare Net, SolarFlareNet, DAFFS, LLMFlareNet, and SwinTCN-Att—depend on photospheric magnetogram data from SDO/HMI, which has been the sole source of high-cadence, full-disk vector magnetograms since its launch in 2010. SDO is currently operating beyond its design lifetime, and no replacement mission with equivalent magnetogram capability has been confirmed beyond ~ 2030 . When SDO ceases operation, every magnetogram-dependent forecasting system will lose its primary input.

GOES XRS data, by contrast, are provided by an operational satellite constellation with guaranteed con-

tinuity through the GOES-U mission (planned through the 2030s) and successors. Multiple GOES spacecraft operate simultaneously, providing built-in redundancy. UEPI-R’s exclusive reliance on XRS data thus provides a forecasting capability that is robust against the single largest instrument-continuity risk facing the space weather community. This makes XRS-only approaches particularly valuable as either standalone systems or as fallback components in ensemble architectures.

C. Comparison with the “No-Skill” Baseline

A critical question for any flare forecasting system is whether it provides skill above simple persistence or climatological baselines [15]. UEPI-R’s false alert rate of $\sim 0.37 \text{ day}^{-1}$ combined with 64–71% coverage is substantially better than a naive threshold detector applied to the raw XRS flux (which would either produce far more false alerts at comparable coverage, or far lower coverage at comparable false rates). The precision-optimized configuration achieves an F_2 score of 0.569 under 1:1 matching, confirming meaningful skill.

Moreover, the C-class analysis of Section IVF demonstrates that the system’s effective false-positive rate is far lower than the M/X-only metric suggests: only 4.2% of alerts occur during periods with no cataloged flare activity at any class. The remaining 95.8% of “false” alerts are detecting real C-class activity—the solar corona is genuinely in an elevated state, but the energy release does not reach M-class. This distinction is operationally important: C-class flares affect HF radio propagation and contribute to cumulative radiation exposure, making many of these alerts actionable even when M-class thresholds are not reached.

D. Future Directions

Several avenues for improvement are under investigation:

1. **Multi-flare matching.** Allowing alerts to claim credit for multiple flares during storm periods would better reflect operational utility but requires careful metric design to avoid inflation.
2. **Probabilistic output.** Converting the binary alert into a continuous probability would enable direct comparison with SWPC forecasts and facilitate downstream decision-making.
3. **EUV augmentation.** Incorporating GOES/SUVI or SDO/AIA EUV data could improve specificity during multi-region activity, while preserving the system’s operational robustness by treating spatial data as optional.

4. **Extended live validation.** Continued operation through Solar Cycle 25 maximum will provide statistically robust live performance estimates.

VII. CONCLUSION

We have presented UEPI-R, a real-time early warning system for M- and X-class solar flares that operates exclusively on GOES XRS irradiance data. Using causal regime detection with physics-motivated diagnostics and no spatial solar data, the system achieves 64–71% M/X coverage and 97.2% X-class coverage over a 16-year validation period (2010–2025), with a day-level TSS of 0.41, a median lead time of 2.4 hours (80% of detections providing ≥ 1 hour warning), precision of $\sim 39\%$, and a false alert rate of $\sim 0.37 \text{ day}^{-1}$.

These results demonstrate that the GOES XRS flux stream contains substantial predictive information for major solar flares—information that is largely unexploited by existing operational systems. Analysis of nominally false alerts reveals that 95.8% are associated with C-class flare activity, with only 4.2% representing true noise triggers—the system detects genuine solar instability in nearly all cases. UEPI-R’s minimal data requirements, strict causality, and real-time capability make it well-suited as a complement to existing magnetogram-based forecasting systems, providing continuous onset warnings where current systems provide only daily probabilities.

The near-perfect X-class detection rate (97.2%) is particularly noteworthy. For the highest-impact space weather events—those most likely to affect satellite operations, aviation, and power grids—UEPI-R provides reliable advance warning using only the most widely available and redundant solar monitoring data.

Data Availability

The GOES XRS data used in this study are publicly available from NOAA’s National Centers for Environmental Information (<https://www.ncei.noaa.gov/data/goes-space-environment-monitor/>) and the SWPC real-time archives (<https://services.swpc.noaa.gov/json/goes/primary/>). The NOAA flare event lists are available from <https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/>. The UEPI-R algorithm is proprietary (U.S. Provisional Patent Application No. 63/949,419); live alerts and verification logs are published at <https://github.com/quantexenergy/UEPI-R-solar-feed>.

ACKNOWLEDGMENTS

The author acknowledges the NOAA Space Weather Prediction Center and the NOAA National Centers for Environmental Information for maintaining and distributing the GOES XRS data products that make this work possible.

-
- [1] A. O. Benz, Flare observations, *Living Reviews in Solar Physics* **14**, 2 (2017).
- [2] L. Fletcher, B. R. Dennis, H. S. Hudson, S. Krucker, K. Phillips, A. Veronig, M. Battaglia, L. Bone, A. Caspi, Q. Chen, P. Gallagher, P. T. Grigis, H. Ji, W. Liu, R. O. Milligan, and M. Temmer, An observational overview of solar flares, *Space Science Reviews* **159**, 19 (2011).
- [3] M. D. Crown, Validation of the NOAA space weather prediction center's solar flare forecasting look-up table and Climatological rates, *Space Weather* **10**, S06006 (2012).
- [4] E. Camporeale and T. E. Berger, Verification of the NOAA space weather prediction center solar flare forecast (1998–2024), *Space Weather* **23**, e2025SW004546 (2025).
- [5] P. S. McIntosh, The classification of sunspot groups, *Solar Physics* **125**, 251 (1990).
- [6] C. J. Schrijver, A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting, *The Astrophysical Journal Letters* **655**, L117 (2007).
- [7] M. G. Bobra and S. Couvidat, Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm, *The Astrophysical Journal* **798**, 135 (2015).
- [8] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, and M. Ishii, Operational solar flare prediction model using Deep Flare Net, *Earth, Planets and Space* **73**, 64 (2021).
- [9] Y. Abdullaah, J. T. L. Wang, Y. Nie, C. Liu, and H. Wang, SolarFlareNet: A deep neural network for solar flare prediction, *Scientific Reports* **13**, 16978 (2023).
- [10] M. K. Georgoulis, D. S. Bloomfield, M. Piana, A. M. Massone, M. Soldati, P. T. Gallagher, E. Pariat, N. Vilmer, E. Buchlin, S. Giordano, and S. Krucker, The flare likelihood and region eruption forecasting (FLARE-CAST) project: Flare forecasting in the big data & machine learning era, *Journal of Space Weather and Space Climate* **11**, 39 (2021).
- [11] S. Riggi, P. Romano, A. Pilzer, and U. Becciani, Solar flare forecasting with foundational transformer models across image, video, and time-series modalities, *Astronomy and Computing* **50**, 100897 (2025).
- [12] Y. Kubo, M. Den, and M. Ishii, Verification of operational solar flare forecast: Case of Regional Warning Center Japan, *Journal of Space Weather and Space Climate* **7**, A20 (2017).
- [13] G. Barnes, K. D. Leka, C. J. Schrijver, T. Colak, R. Qahwaji, O. W. Ashamari, Y. Yuan, J. Zhang, R. T. J. McAteer, D. S. Bloomfield, P. A. Higgins, P. T. Gallagher, D. A. Falconer, M. K. Georgoulis, M. S. Wheatland, C. Balch, T. Dunn, and E. L. Wagner, A comparison of flare forecasting methods. I. Results from the “All-Clear” workshop benchmarks, *The Astrophysical Journal* **829**, 89 (2016).
- [14] K. D. Leka, S.-H. Park, K. Kusano, J. Andries, G. Barnes, S. Bingham, D. S. Bloomfield, A. E. McCloskey, P. T. Gallagher, D. A. Falconer, J. Jing, J. Lee, C. Liu, H. Wang, and K. D. Leka, A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems, *The Astrophysical Journal Supplement Series* **243**, 36 (2019).
- [15] D. S. Bloomfield, P. A. Higgins, R. T. J. McAteer, and P. T. Gallagher, Toward reliable benchmarking of solar flare forecasting methods, *The Astrophysical Journal Letters* **747**, L41 (2012).
- [16] K. D. Leka, G. Barnes, and E. L. Wagner, The NWRA classification infrastructure: Description and extension to the Discriminant Analysis Flare Forecasting System (DAFFS), *Journal of Space Weather and Space Climate* **8**, A25 (2018).
- [17] Y. Zhang, B. Liang, and S. Feng, Solar flare forecasting based on swin transformer and temporal convolutional networks, *Astrophysics and Space Science* **370**, 98 (2025).
- [18] X. Li, Y. Lv, J. Wei, Y. Zheng, T. Li, R. Wang, Z. Wu, H. Ye, P. Yan, Z. Zainal Abidin, N. Mohamed Shah, C. Xiang, S. Zhang, X. Ji, X. Huang, X. Wang, and H. Jin, Operational solar flare forecasting system using an explainable large language model, arXiv e-prints (2026), [arXiv:2601.22811](https://arxiv.org/abs/2601.22811).