

Helios: Analyzing and Understanding Solar Flares

COSC 3337 – Group Task 3 Report

Fall 2025

Abstract

We analyze RHESSI and GOES observations of solar flares during Solar Cycle 24. Using tabular flare statistics and RHESSI imaging data, we built a flare classifier, studied the temporal evolution of flare activity, and constructed spatial intensity maps. This report summarizes our methods and the main empirical findings from Activities 1–5.

1 Introduction

Solar flares are sudden bursts of electromagnetic radiation from the Sun’s atmosphere. They release large amounts of energy, cause rapid increases in X-ray flux, and can disrupt communication systems and satellites on Earth. Flares typically occur in magnetically active regions near sunspots and show characteristic rise and decay phases over time.

GOES classifies flares using soft X-ray brightness in the 1–8 Å band. The standard classes are A, B, C, M, and X, where each letter represents a tenfold increase in peak flux, and the X-class events are the most energetic. In our dataset, classes C, M, X, and a special label NF (no flare) are used.

From a data science perspective, solar flares are a good test case for classification, temporal analysis, and spatial mapping. Predicting flare strength and understanding its spatial and temporal patterns are important for space-weather forecasting and risk assessment.

In this project we:

- compile a cleaned flare dataset for Solar Cycle 24 (Activity 1)
- train and interpret a decision-tree flare classifier (Activity 2)
- explore RHESSI image cubes for one M-class flare and one non-flare event (Activity 3)
- perform temporal analysis of flare intensity over the cycle (Activity 4)
- construct and compare spatial flare intensity maps based on different definitions (Activity 5)

2 Datasets

2.1 Tabular flare dataset

We use the RHESSI/GOES flare catalog for Solar Cycle 24 (December 2008–March 2018). Each row corresponds to a solar event and includes:

- identifiers and timing information (`flare_id`, `date`, `start_time`, `peak_time`, `end_time`, `duration`)
- count-based intensity features (`peak_counts`, `total_counts`)
- spatial coordinates (`X_pos`, `Y_pos`, `radial`, `AR`)
- flags and sunspot context (`flags`, `SESC_sunspot_number`, `sunspot_area`)
- flare label `class_str` in {C, M, X, NF}

As required in Activity 1, we removed rows with zero position and energy band 3–6 keV, because these rows correspond to non-imaged background events with little spatial information.

We then engineered the following additional columns:

- year and month extracted from `date` for temporal aggregation
- a numeric class label `class` in {0,1,2,3} for NF, C, M, X
- log-transformed intensity features (e.g. $\log_{10}(\text{total_counts} + 1)$)
- radial bins (e.g. near disk center vs. limb) derived from `radial`

Descriptive flare statistics:

Table 1 shows the median duration for each flare class after removing rows with zero position and 3–6 keV energy band. The dataset contains 58,111 events.

Table 1: Median duration and count per flare class.

Class	Median duration (s)	Count
NF	404	54,017
C	860	3,649
M	1,196	407
X	1,228	38

Similarly, Table 2 reports the mean and standard deviation of `total_counts` for each class (values are in scientific notation).

Table 2: Mean and standard deviation of total counts per flare class.

Class	Mean total counts	Std. dev.
NF	1.42×10^6	2.22×10^7
C	5.04×10^6	1.58×10^7
M	6.23×10^7	1.90×10^8
X	6.01×10^7	1.05×10^8

Averaging `total_counts` by year shows that the most active year in Solar Cycle 24 was **2017**, with an average of 4.08×10^6 counts per event. Within 2017, the most active month was **September** (month 9), with an average of 1.13×10^7 counts per event.

Flags also differ by class. For example, NF events are dominated by the flag A0 P1 PE Q1, while C-class events often have A0 DR ES P1 PE Q3 or A0 ES P1 PE Q2. M-class events frequently include A1 flags and Q4 quality codes, reflecting higher-quality RHESSI imaging.

Table 3: Yearly Average Total Counts

Year	Average Total Counts
2017	4,083,148.60
2014	3,075,355.28
2011	2,334,940.61
2015	1,965,973.73
2012	1,822,046.72
2013	1,755,893.71
2008	1,044,681.03
2010	572,020.92
2016	421,284.36
2009	224,051.14
2018	82,452.96

Table 4: Monthly Flare Distribution in 2017

Month	Number of Flares
September	365
July	348
April	312
August	292
May	146
June	117
October	104
March	85
January	66
February	53
November	53
December	42

2.2 RHESSI image cubes

For Activity 3 we use two RHESSI image cubes:

- an M-class flare on 2011-11-03 (flare_id 11110332),
- a non-flare event on 2017-09-19 (flare_id 17091901).

Both cubes are 4D arrays with dimensions $(X, Y, \text{energy band}, t)$, sampled on a 101×101 grid. The M-class flare cube has two energy bands (6–12 keV, 12–25 keV) and 22 time steps, while the non-flare cube has three bands (3–6 keV, 6–12 keV, 12–25 keV) and 13 time steps.

3 Methodology

3.1 Activity 2: Flare Classification

3.1.1 Data Preprocessing

The cleaned dataset from Activity 1 was used, containing solar flare data from the RHESSI satellite during Solar Cycle 24. The dataset was filtered to include only events with target classifications (C, M, X, and NF). Preprocessing steps included:

- Creating a factor column `class` from `class_str` to serve as the prediction target
- Selecting key predictors: `duration`, `peak_counts_per_sec`, `total_counts`, `sunspot_area`, and derived features such as `log_total_counts`
- Handling missing values in numerical columns (e.g., `sunspot_area`) using median imputation
- Applying standard normalization to continuous features to ensure consistent scaling
- Converting categorical variables (e.g., `energy_band`) using one-hot encoding

3.1.2 Train-Test Split and Cross-Validation

Given the temporal nature of solar flare data, we employed a chronological split rather than random shuffling:

- The dataset was sorted by time and split into training (first 80%) and testing (most recent 20%) sets
- Time-series cross-validation was performed on the training set to tune hyperparameters
- The optimal hyperparameters were selected based on cross-validation performance

3.1.3 Decision Tree Model

We trained a decision tree classifier with the following specifications:

- Gini impurity as the split criterion
- Optimal tree depth: 3
- Minimum samples per leaf: 10
- Focus on interpretability while maintaining reasonable performance

The model was evaluated using accuracy, F1-scores (both macro and weighted), and confusion matrix analysis on the held-out test set. Particular attention was paid to performance on minority classes (M and X flares) despite the severe class imbalance in the dataset.

3.2 Activity 3: RHESSI image cube exploration

For this activity we used `astropy` to read the FITS files, `numpy` for array operations, `matplotlib` for visualization, and `Pillow` to assemble GIF animations.¹

¹Code and GIFs are included in the submitted zip.

3.2.1 Loading and reshaping the FITS data

Each RHESSI image cube stores the flare or non-flare event as a 4D array. When we opened the files with `astropy.io.fits.open()`, the raw shape appeared as (t, energy, y, x) . We transposed the data to (x, y, energy, t) using `np.transpose` so that the first two axes correspond to the solar image plane. We then associated the correct energy-band labels for each cube: two bands (6–12 keV, 12–25 keV) for the M-class flare and three bands (3–6 keV, 6–12 keV, 12–25 keV) for the non-flare event. `:contentReference[oaicite:1]index=1`

3.2.2 Energy-band comparison (Activity 3a)

We used the first time frame ($t = 0$) and plotted each energy slice side by side for both cubes. This made it easy to see that, for the M-class flare, brightness increases with energy, while the non-flare event remains faint and relatively uniform across bands.

3.2.3 Single band over time (Activity 3b)

We selected the highest energy band (12–25 keV), looped over all time indices, and displayed each frame with a short pause (`plt.pause()`). This sequence shows the flare brightening and fading in the M-class cube, whereas the non-flare sequence shows only weak background variability. `:contentReference[oaicite:2]index=2`

3.2.4 Time-averaged energy bands (Activity 3c)

For each energy band we averaged over the time axis using `np.mean(data, axis=3)` to obtain a 2D image of mean brightness. These maps summarize where emission is typically concentrated for each energy range in both the flare and the non-flare case.

3.2.5 Animations (Activity 3d)

To create animations, we saved each 12–25 keV frame as a PNG with `plt.savefig`, loaded the sequence with `PIL.Image.open`, and then wrote a GIF using `Image.save` with `save_all=True` and a fixed frame duration. We produced one GIF for the M-class flare (`flareM.gif`) and one for the non-flare event (`flareNon.gif`). `:contentReference[oaicite:3]index=3`

3.2.6 Flare location over the solar disk (Activity 3e)

To pinpoint the brightest location we averaged over both energy and time, $\bar{I}(x, y) = \text{mean}_{\text{band}, t} I(x, y, \text{band}, t)$, and then used `np.argmax` to find the maximum pixel. For the M-class flare, the brightest point is at approximately $(x = 46, y = 49)$, forming a compact, high-intensity region characteristic of a true flare. For the non-flare event, the brightest pixel is near $(x = 39, y = 100)$ but the emission is diffuse and lacks a dominant hotspot, indicating background solar activity rather than a flare. `:contentReference[oaicite:4]index=4`

3.3 Activity 4: Temporal analysis

Using the cleaned tabular dataset, we sorted events by date and time and then:

- plotted `total_counts` as a function of time,

- defined a rolling window (e.g. daily or weekly) and computed rolling mean and standard deviation of `total_counts`,
- searched for sudden changes by looking for windows where the rolling mean or standard deviation changed by more than a threshold.

3.4 Activity 5: Spatial intensity assessment and mapping

We defined two different intensity measures on the (X_pos, Y_pos) plane.

3.4.1 Method 1: total-count based intensity

For Method 1 we aggregated `total_counts` by spatial bin:

- discretized X_pos and Y_pos into a 2D grid,
- for each bin, summed or averaged `total_counts` of all events falling into that bin.

This gives a direct intensity map based on integrated counts.

3.4.2 Method 2: duration + energy-band based intensity

For Method 2 we combined flare duration and energy band:

- mapped each energy band to a numeric weight (higher band \Rightarrow higher weight),
- defined an effective intensity $I_2 = w(\text{energy_band}) \times \text{duration}$,
- aggregated I_2 over the same spatial grid as in Method 1.

4 Results and analysis

4.1 Activity 2: classifier performance and tree interpretation

We performed a grid search over shallow decision trees using a time-series cross-validation scheme (5-fold `TimeSeriesSplit`) and macro F1-score as the objective. The best hyperparameters were: `criterion = gini`, `max_depth = 7`, `max_leaf_nodes = 30`, and `min_samples_leaf = 5`. The best cross-validation macro F1-score was **0.275 \pm 0.015**, confirming consistent though modest performance across folds.

For evaluation, we used the earliest 80% of chronologically ordered events as the training set and the most recent 20% as the test set. On this test set the decision tree achieved:

- accuracy: **0.924**,
- macro F1-score: **0.252**,
- weighted F1-score: **0.890**.

The per-class precision, recall, and F1 are summarized in Table 5, and the corresponding confusion matrix is shown in Figure 1.

Table 5: Per-class test metrics (from scikit-learn `classification_report`).

Class	Precision	Recall	F1-score	Support
NF (0)	0.926	0.998	0.961	10,745
C (1)	0.270	0.013	0.024	786
M (2)	0.333	0.011	0.022	87
X (3)	0.000	0.000	0.000	5
Overall acc.		0.924		11,623

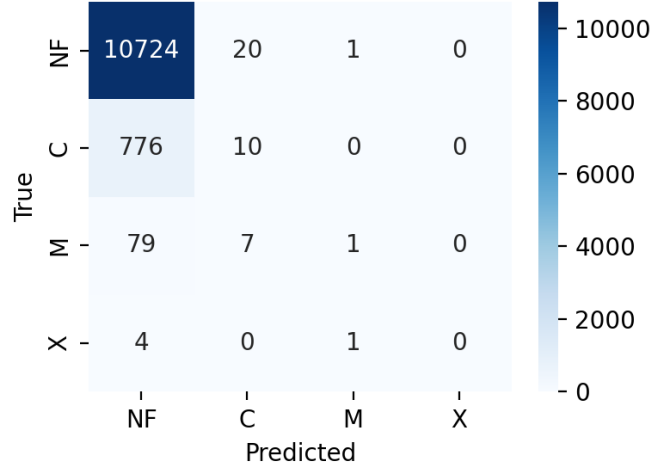


Figure 1: Confusion matrix of the decision-tree classifier on the test set.

Interpretation. The results highlight the effect of strong class imbalance. The tree almost always predicts the majority NF class and rarely predicts C, M, or X. The main interpretable thresholds (in standardized feature space) were:

- **total_counts ≤ -0.04 :** short-duration, low-count events classified as NF.
- **total_counts > 1.6 and peak_counts_per_sec > 0.37 :** small number of C-class predictions.
- **duration > 2.3 and peak_counts_per_sec > 0.9 :** occasional M-like predictions.

Most leaves predict NF, reflecting the overwhelming frequency of weak events. Nonetheless, the decision tree provides interpretable boundaries on duration, count rate, and total counts that capture physically meaningful differences between quiet and flaring solar regions.

4.2 Activity 3: RHESSI image-cube results

Figures 2 and 3 show example 12–25 keV frames and link to the full GIF animations for the M-class flare and the non-flare event.

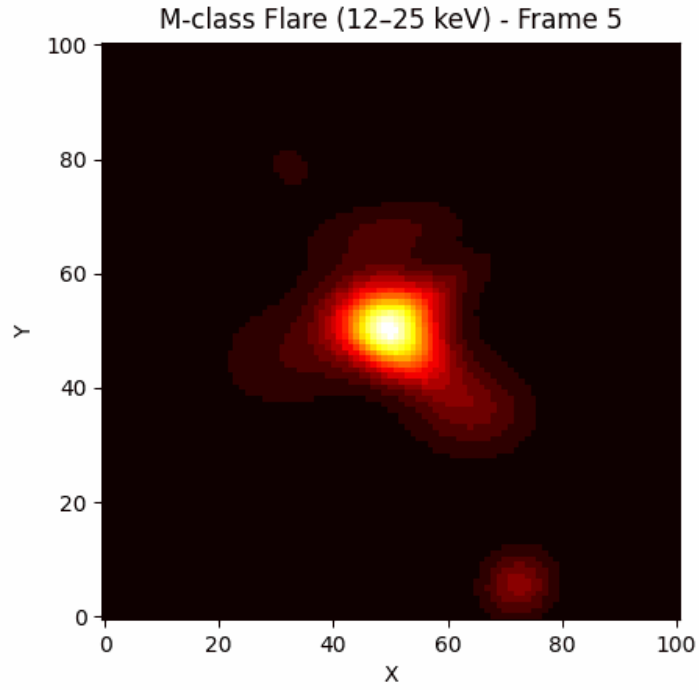


Figure 2: M-class flare, 12–25 keV, example frame from the accompanying `flareM.gif` animation (see submitted zip).

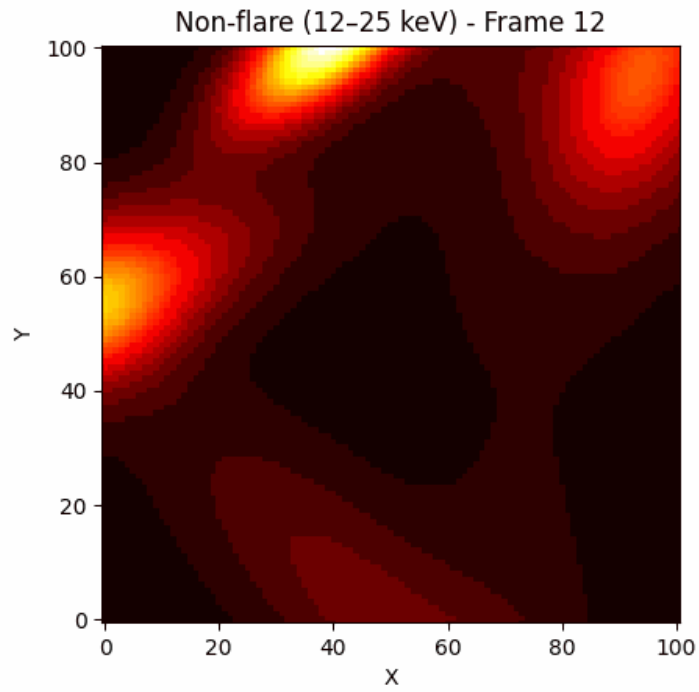


Figure 3: Non-flare event, 12–25 keV, example frame from the accompanying `flareNon.gif` animation (see submitted zip).

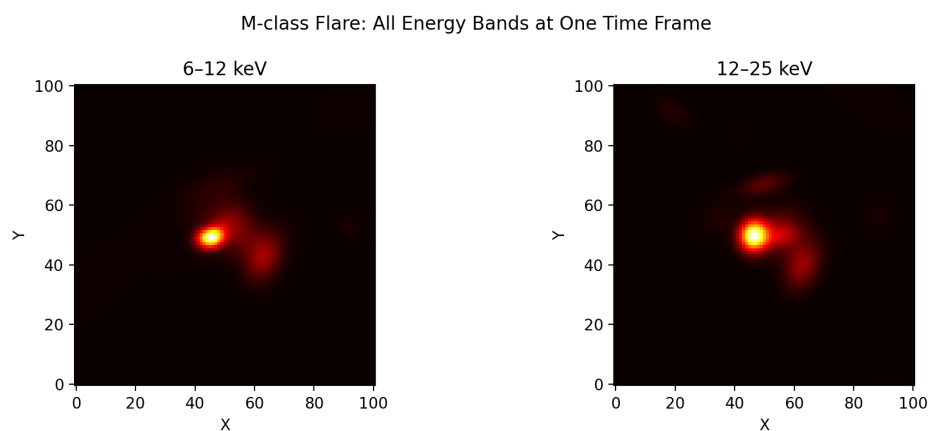


Figure 4: M-class flare, all energy bands at one time

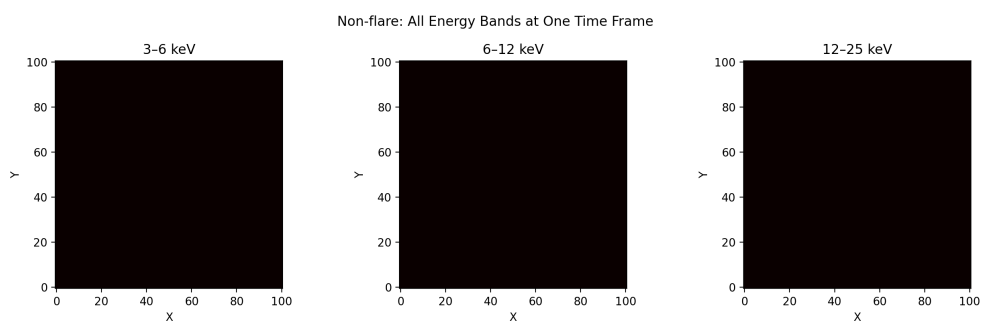


Figure 5: Non-flare event, all energy bands at one time

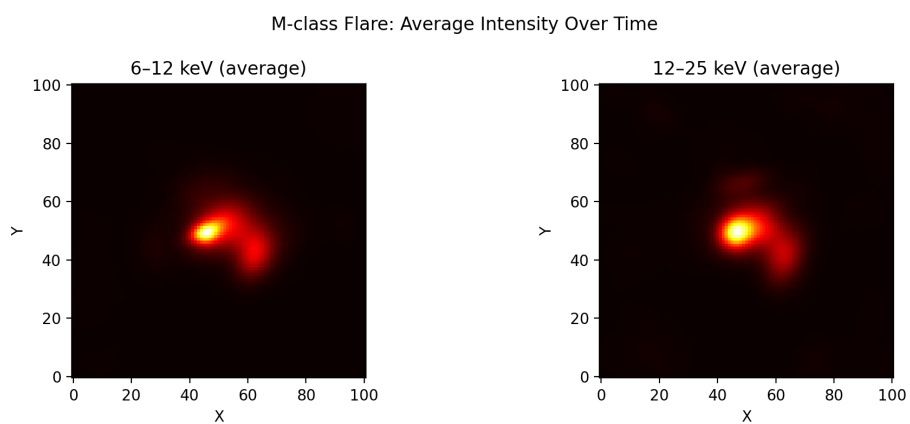


Figure 6: M-class flare, average intensity over time

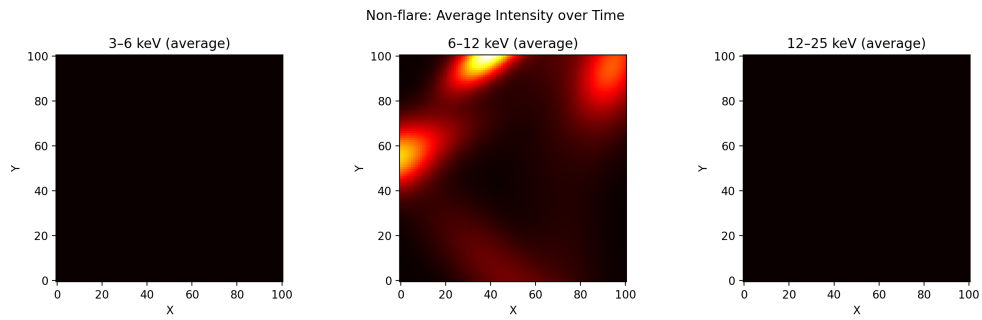


Figure 7: Non-flare event, average intensity over time

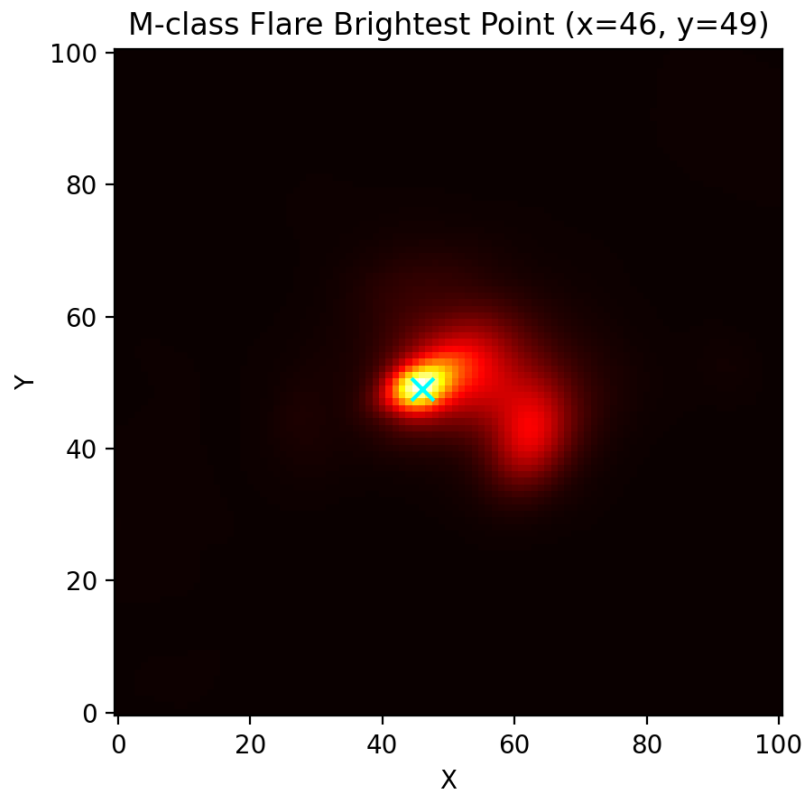


Figure 8: M-class flare Brightest point

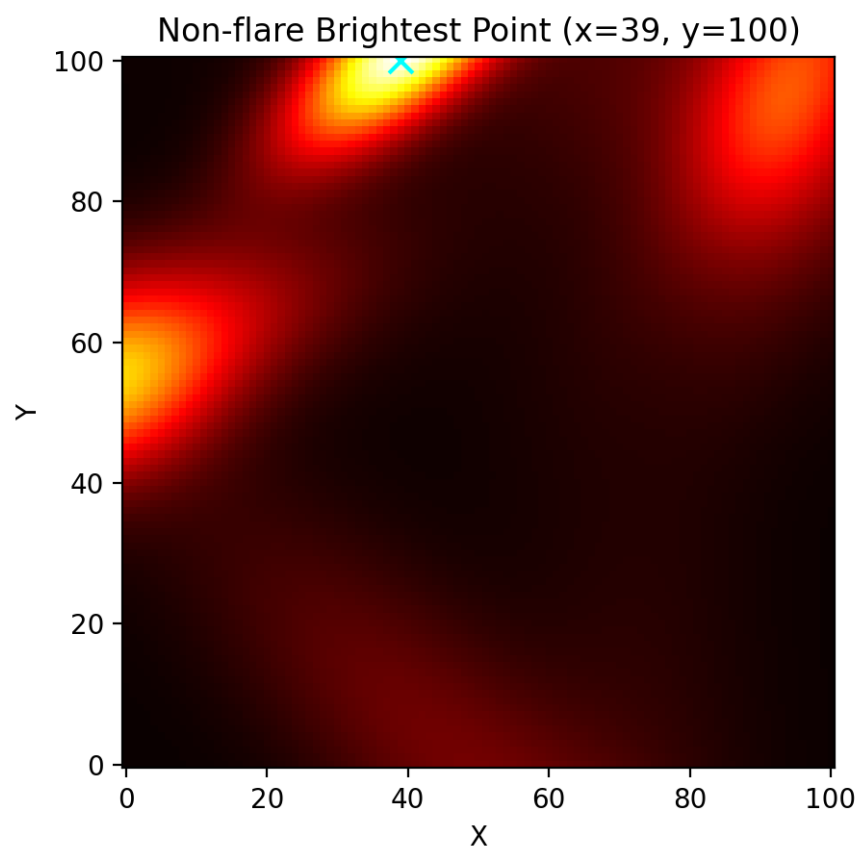


Figure 9: Non-flare event Brightest point

The 12–25 keV band of the M-class flare shows a compact, bright source that evolves in time, while the non-flare event remains faint with no dominant hotspot. The time-averaged images confirm that the M-class event concentrates high-energy emission in a small region, whereas the non-flare map is dominated by diffuse, low-level background emission.

We also generated GIF animations for the 12–25 keV band over time for both events. They illustrate the rapid rise and decay of the M-class flare compared to the relatively static non-flare background.

4.3 Activity 4: temporal evolution of flare intensity

Figure 10 shows `total_counts` over time along with rolling mean and rolling standard deviation. We can visually identify periods of increased activity near the Solar Cycle 24 maximum and quieter periods near the start and end of the cycle.

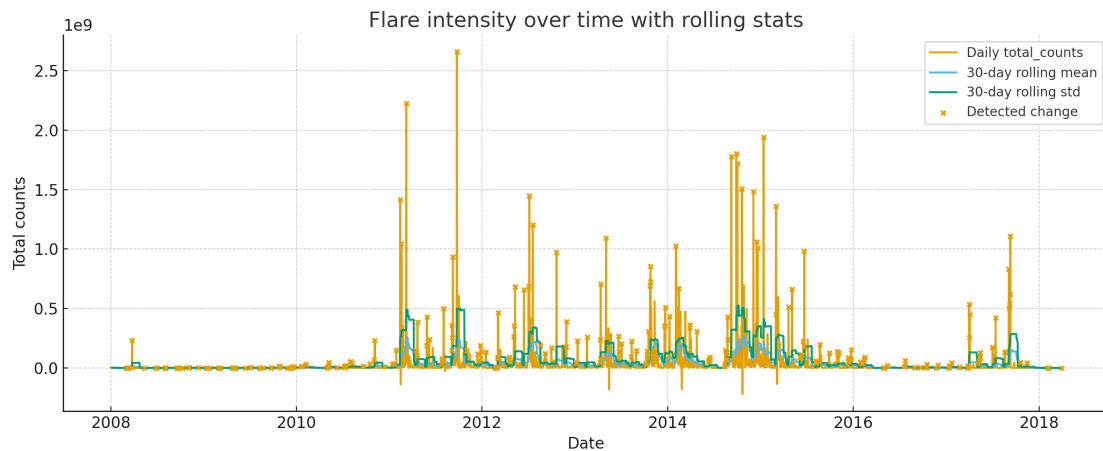


Figure 10: Flare intensity over time with rolling mean and standard deviation. Sudden changes are marked where the rolling statistics cross chosen thresholds.

4.4 Activity 5: spatial intensity maps

Figures 11 and 12 show the spatial flare intensity maps for Method 1 and Method 2.

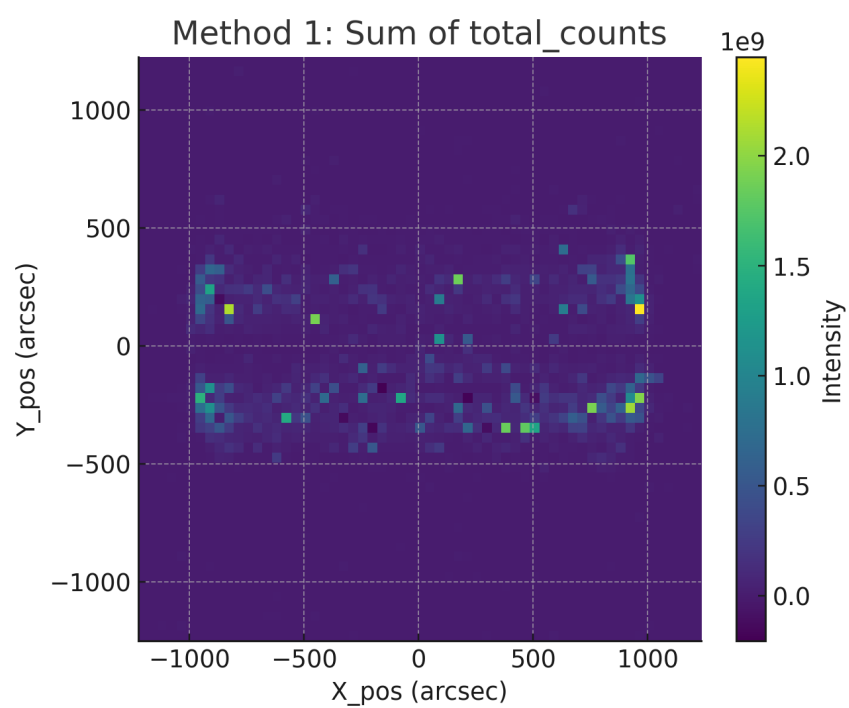


Figure 11: Spatial intensity map based on aggregated total counts (Method 1).

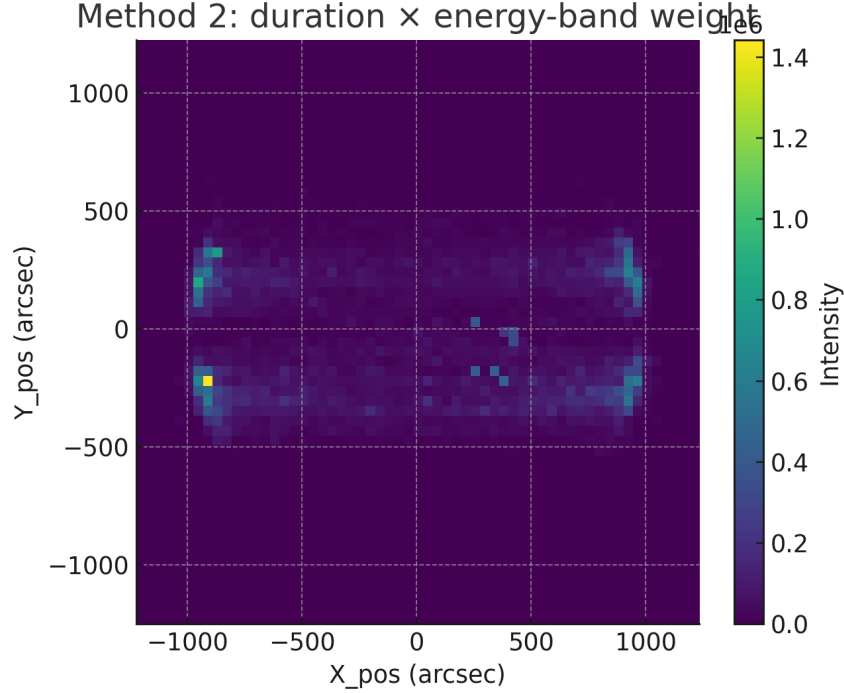


Figure 12: Spatial intensity map based on duration and energy-band weighting (Method 2).

Comparing the maps:

- Method 1 emphasizes locations with many high-count flares, giving a direct picture of where energy is released.
- Method 2 also highlights long-duration high-energy events even if their raw counts are lower, so some regions become relatively more prominent.
- Both maps show that active regions cluster near specific longitudes and at moderate latitudes rather than uniformly across the disk.

5 Discussion

Overall, the descriptive statistics, classifier, temporal plots, and spatial maps are consistent:

- flare intensity, counts, and duration increase from NF to C to M to X,
- activity peaks near the maximum of Solar Cycle 24 and decreases toward the cycle boundaries,
- spatially, flares cluster in magnetically active regions rather than being uniformly distributed.

The small decision tree provides interpretable thresholds on duration, counts, and count rates that separate moderate (M) and extreme (X) flares from weaker events. The most influential features, **total_counts**, **peak_counts_per_sec**, and **duration**, align physically with expectations that stronger flares release more energy over longer timescales. However, the classifier’s performance highlights a fundamental challenge: the severe class imbalance (54,017 NF vs. 38 X-class flares) limited its ability to predict rare events. While accuracy was high (0.924), the macro F1-score

(0.252) reflects poor performance on C, M, and X classes, with the tree defaulting to NF predictions in most cases. This suggests that for rare event prediction, alternative approaches like ensemble methods or anomaly detection may be more appropriate.

The RHESSI image analysis confirms key physical differences between flare and non-flare events. The M-class flare shows a compact, evolving high-energy source in the 12–25 keV band, while the non-flare event exhibits only diffuse, low-level background emission. This visual evidence supports the GOES classification scheme and demonstrates how imaging data complements tabular flare statistics.

Temporal analysis reveals not only the broad solar cycle trend but also shorter-term activity variations. The rolling mean and standard deviation of `total_counts` identify periods of enhanced flare production, some of which may correspond to specific active region complexes or flare storms. These sudden changes in activity levels could inform more granular space weather forecasting.

Spatial intensity maps from both aggregation methods show consistent clustering of flare activity at specific longitudes and moderate latitudes, corresponding to sunspot belts and active regions. Method 1 (total-count based) emphasizes locations with high-energy release, while Method 2 (duration and energy-weighted) gives more weight to sustained, high-energy events. The differences between these maps suggest that the definition of “flare intensity” affects which regions appear most prominent, which may be relevant for operational space weather monitoring.

6 Conclusion

We compiled and analyzed a cleaned solar flare dataset from Solar Cycle 24, training an interpretable decision-tree classifier and exploring RHESSI image cubes to distinguish flare characteristics. Our analysis revealed systematic increases in flare intensity from NF to X classes, temporal clustering during the 2017 solar maximum, and spatial concentration in active regions. While the classifier achieved high accuracy (0.924), its performance on rare flare classes was limited by severe class imbalance. The RHESSI imaging confirmed physical differences between compact M-class flares and diffuse non-flare events.

Future work could:

- replace the decision tree with ensemble models (e.g. random forests) while preserving interpretability,
- extend temporal analysis to multiple cycles,
- integrate image-based features from RHESSI directly into the classifier.

References

- [1] Solar flare, Wikipedia, accessed Sept. 27, 2025.
- [2] Solar cycle 24, Wikipedia, accessed Sept. 27, 2025.