Large Scale Evaluation of Deep Learning-based Explainable Solar Flare Forecasting Models with Attribution-based Proximity Analysis

Temitope Adeyeha, Chetraj Pandey, Berkay Aydin

Dept. of Computer Science, Georgia State University, Atlanta, US

{tadeyeha1, cpandey1, baydin2}@gsu.edu

Abstract—Accurate and reliable predictions of solar flares are essential due to their potentially significant impact on Earth and space-based infrastructure. Although deep learning models have shown notable predictive capabilities in this domain, current evaluations often focus on accuracy while neglecting interpretability and reliability—factors that are especially critical in operational settings. To address this gap, we propose a novel proximity-based framework for analyzing post hoc explanations to assess the interpretability of deep learning models for solar flare prediction. Our study compares two models trained on full-disk line-of-sight (LoS) magnetogram images to predict >M-class solar flares within a 24-hour window. We employ the Guided Gradient-weighted Class Activation Mapping (Guided Grad-CAM) method to generate attribution maps from these models, which we then analyze to gain insights into their decision-making processes. To support the evaluation of explanations in operational systems, we introduce a proximity-based metric that quantitatively assesses the accuracy and relevance of local explanations when regions of interest are known. Our findings indicate that the models' predictions align with active region characteristics to varying degrees, offering valuable insights into their behavior. This framework enhances the evaluation of model interpretability in solar flare forecasting and supports the development of more transparent and reliable operational systems.

Index Terms—Explainable AI, Deep Learning, Solar Flares

I. INTRODUCTION

Solar flares are transient events characterized by the rapid release of electromagnetic radiation in the outer layers of the Sun's atmosphere. These phenomena play a central role in space weather forecasting and can significantly impact various stakeholders, including power grids, satellite communications, and aviation. Flares are classified into five categories—A, B, C, M, and X—each further subdivided into a strength rating ranging from 1.0 to 9.9, based on their peak X-ray flux, as outlined by the National Oceanic and Atmospheric Administration (NOAA) [1]. M- and X-class flares are particularly important to predict due to their potential to cause substantial disruptions on Earth and in space [2].

Active Regions (ARs) on the Sun, distinguished by disturbed magnetic fields, are the primary initiators of solar flares and other space weather events [3]. Many solar flare prediction models analyze data across the entire solar disk (full-disk models [4]–[6]), which can make it challenging to pinpoint the relevant ARs responsible for flares. These models, often neural networks, are referred to as "opaque" or "black boxes" due to their complex learning representations, which obscures

their prediction rationale. This lack of transparency raises concerns for operational forecasting systems, highlighting the need to ensure that these models are interpretable and their predictions understandable for successful integration into real-world forecasting applications. To address these challenges, empirical methods from broader machine learning research have been adopted to better understand how neural networks make decisions. Post hoc explanations, often referred to as attribution methods (e.g., [7]–[10]), aim to interpret model outputs by identifying which input features contributed most to a given prediction. These methods enhance transparency by providing insights into the internal workings of complex models, making them more trustworthy for critical applications. However, challenges remain in quantitatively evaluating the effectiveness and reliability of these attribution methods, which is crucial for ensuring they provide meaningful and actionable explanations in operational flare forecasting.

Recently, several studies [6], [11]–[13] have focused on developing explainable models for solar flare prediction. To interpret a convolutional neural network (CNN) trained on AR patches, Bhattacharjee et al. [11] employed an occlusion-based method. More recently, [14] evaluated DeepLIFT [15] and Integrated Gradients [9] for CNNs trained on central AR patches (within ±70°). Pandey et al. [12], [16] extended this work by employing explainable full-disk flare prediction models and qualitatively analyzing near-limb flare predictions. This was further expanded in [6] with a questionnaire-based evaluation of explanations for X-class flares, demonstrating the model's effectiveness across both near-limb and central regions and validating its feasibility for operational forecasting. While some studies emphasize model development [4], [17], [18], others incorporate explainability techniques [6], [14], [19]; however, none offer automated evaluation of explanations.

Building on our prior work [6], [18], which introduced an explainable full-disk solar flare prediction model using compressed line-of-sight (LoS) magnetograms and evaluated Guided Grad-CAM (GGCAM) [7] explanations with human-in-the-loop evaluations in [6], this paper advances the field with a fully automated system for evaluating explanations. Designed to localize predictions to ARs from full-disk models, this system enhances automation, marking a significant step toward improving solar flare forecasting and increasing trust in predictive models.

The primary contributions of this paper are: (i) We introduce a fully automated quantitative approach for analyzing model explanations, (ii) We introduce novel evaluation metrics to define distance-based relationships between model explanations, when the regions of interest are known, and (iii) we present a comprehensive case study comparing the explanations of two standard CNN-based models from [6] and [18], using proximity score and colocation ratio.

The remainder of this paper is structured as follows: Sec. II details our methodology, including data preparation and the proximity-based approach of evaluating explanations. Sec. III presents our experimental evaluation and results. Sec. IV discusses the analysis and the implications of our findings for the validity of deep learning models. Finally, Sec. V concludes the paper and outlines directions for future research.

II. METHODOLOGY

In this section, we describe our approach for evaluating explanations. We utilized two full-disk deep learning models presented in [6] and [18] to assess the efficacy of our method. These CNN-based models were trained using full-disk line-of-sight (LoS) magnetograms from 2010 to 2018 to predict $\geq$M-class flares, with a prediction window of 24 hours. Details on the model configurations and their predictive performance can be found in the relevant papers cited above.

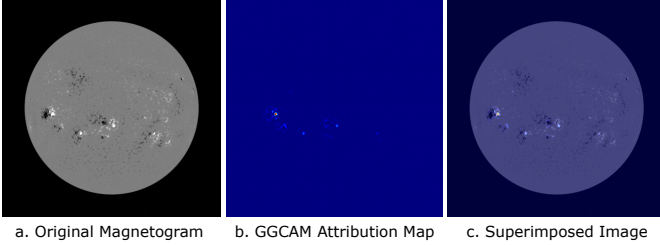a. Original Magnetogram   b. GGCAM Attribution Map   c. Superimposed Image

Fig. 1. (a) Original input magnetogram image at 2015-01-01T07:59:39.30 UTC. (b) Attribution map generated by Guided Grad-CAM (GGCAM) for input (a). (c) Superimposed image of input (a) and attribution map (b).

For this study, we primarily require full-disk magnetogram images from the Helioseismic and Magnetic Imager (HMI) aboard the Solar Dynamics Observatory (SDO), which represent the Sun's magnetic field strength, with white and black areas indicating positive and negative polarities. These images, as shown in Fig. 1 (a), are publicly available on Helioviewer[1]. Furthermore, given an input magnetogram image to the trained full-disk models [6], [18], we generate attribution maps referred to as explanations. Attribution maps are a type of local post hoc explanation that shows the relative importance of pixels for a given prediction, i.e., the attribution of pixels for a particular decision. We used Guided Grad-CAM (GGCAM) [7], a gradient-based method that combines Grad-CAM's localization ability with the pixel-level precision of guided backpropagation [8] to identify important regions in magnetogram images (e.g., see Fig. 1(b)). Furthermore, for proximity analysis of explanations corresponding to actual

---

[1]https://api.helioviewer.org/docs/v2/

flare locations, we used NOAA-detected flare database[2], which include solar coordinates and associated AR numbers. To ensure a clear comparison between the attribution maps and the magnetograms, preprocessing steps are applied to standardize the attribution maps. Both magnetogram and attribution maps are converted into numpy arrays, normalized, and scaled to 255. For visualization purposes, the arrays are superimposed with an alpha saturation of 0.5, as demonstrated in previous studies [6] (Fig. 1 (c)).

Algorithm 1 PEPS-SF: Proximity-based Evaluation of Post-hoc Explanations in Solar Flare Prediction

---

Input: M (Magnetogram), AR (AR locations), FL (FL locations)

Output: Proximity Scores (PS), Attribution Colocation Ratio (ACR)

  Generate Attribution Maps:
1: Attribution Map (AM) $\leftarrow$ GGCAM($model, M$)
  Scaling (0-255):
2: $G \leftarrow$ scale(AM)
  Edge Detection:
3: $P \leftarrow$ edge_detection(G)
  Density-Based Clustering:
4: $C, Clusters \leftarrow$ DBSCAN(P)
  Get Convex Hull and Add Buffer Pixels:
5: **for** each $c \in Clusters$ **do**
6:   $CH \leftarrow$ convex_hull(c)
7:   $BR \leftarrow$ add_buffer(CH)
8: **end for**
  Coordinate Conversion:
9: $AR_{pix}, FL_{pix} \leftarrow$ convert_coordinates(AR, FL)
  Proximity Analysis:
10: **for** each $ar \in AR_{pix}$ **do**
11:   $BR_{closest} \leftarrow$ find_closest(ar, BR)
12:   **if** $ar \in BR_{closest}$ **then**
13:     $d \leftarrow 0$
14:   **else**
15:     $d \leftarrow$ euclidean_distance($ar, BR_{closest}$)
16:   **end if**
17: **end for**
  Calculate Metrics:
18: $PM \leftarrow$ average(d)
19: $ACR \leftarrow \frac{\#AR_{pix} \in BR}{\#AR_{pix}}$
20: **return** PS, ACR

A. Preprocessing Attribution Maps

To quantitatively assess the alignment of highlighted pixels in the attribution maps with observed ARs, we first extract pixel locations and compare them to the AR coordinates. We then apply Canny Edge Detection and density-based clustering to group the important pixels, followed by deriving the convex hull for each group. Next, we compare the coordinates of these convex hulls with the coordinates of the ARs and flares
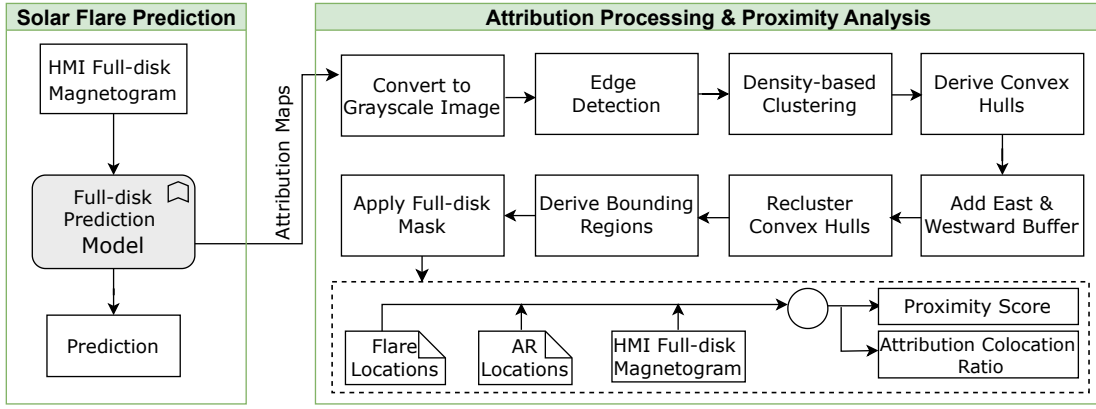
---

[2]ftp://ftp.swpc.noaa.gov/pub/warehouse

Fig. 2. An schematic overview of steps involved in attribution maps preprocessing and corresponding proximity-based analysis used in this study.

in the original image. To ensure consistency in proximity analysis, we convert spatial coordinates to pixel coordinates, establishing a common scale. This conversion also accounts for the Sun's rotation, which causes an approximate westward shift of 13 degrees over a 24-hour period, by introducing buffers to the east and west around each convex hull. The entire procedure for proximity-based solar flare prediction and analysis is detailed in Algorithm 1 (PEPS-SF), which outlines the generation of attribution maps, scaling, edge detection, density-based clustering, convex hull derivation, and the application of full-disk region (circular mask). Additionally, the algorithm includes the conversion of AR and flare location (FL) coordinates to pixel coordinates, followed by a proximity analysis that calculates the Euclidean distance between each AR and the nearest point on the convex hull.

The system for solar flare prediction and analysis is composed of interconnected modules as shown in Fig. 2. The Solar Flare Prediction Module initiates the process by processing full-disk magnetogram images, generating prediction and attribution maps that highlight potential solar flare regions, as described in step 1 of Algorithm 1. These maps are then preprocessed by the Attribution Map Processing Module, which executes steps 2 to 9 in Algorithm 1, resulting in a set of maps ready for detailed analysis. For proximity analysis, we use these preprocessed maps and compares them with observational data, including the locations of AR and FL, to calculate the Proximity Score (PS) and Attribution Colocation Ratio (ACR) using steps 10 to 20 in Algorithm 1. This system design provides a structured framework for evaluating how well the activated pixels in the post hoc explanations correspond with actual solar flare activity.

For location-based analysis, heliographic coordinates are converted to pixel coordinates to facilitate accurate proximity analysis, ensuring all spatial data and pixel locations are on a common scale, thereby enabling precise calculations. This conversion is executed using the transformations provided in Eq. (1) and Eq. (2):

$$\mathrm{pix}_x = \mathrm{HPCCENTER} + \frac{\mathrm{hpcx}}{\mathrm{CDELT}} \qquad (1)$$

$$\mathrm{pix}_y = \mathrm{HPCCENTER} - \frac{\mathrm{hpcy}}{\mathrm{CDELT}} \qquad (2)$$

where pix_x and pix_y represent the pixel coordinates on the image, HPCCENTER denotes the center of the solar disk in pixel coordinates, hpcx and hpcy are the Helioprojective-Cartesian coordinates, and CDELT is the scaling factor (i.e., pixel size in degrees). To measure the proximity of active regions to their corresponding convex hulls, the closest convex hull to each active region is identified by computing the Euclidean distance between the region's coordinates and the nearest point on the convex hull, as described by Eq. (3).

$$d_{\min} = \sqrt{(p_x^* - q_x)^2 + (p_y^* - q_y)^2} \qquad (3)$$

Here, $d_{\min}$ represents the minimum Euclidean distance, with (q_x, q_y) denoting the coordinates of the solar active region and (p_x^*, p_y^*) representing the coordinates of the closest point on the convex hull.

A working example is provided in Fig. 3, which outlines the sequence of transformations applied to an attribution map. The process starts with In **Step (a)**, the Guided-GradCAM image is resized and pixel intensities are scaled to 255. This is followed by edge detection in **Step (b)** and clustering of the detected edges in **Step (c)**. In **Step (d)**, a convex hull is drawn around the clustered points to define the regions of interest. To account for the Sun's rotation, eastward and westward buffers are added in **Step (e)**. In **Step (f)**, these buffered convex hulls are re-clustered to merge smaller clusters into larger, more cohesive clusters. A bounding region is then created around these clusters in **Step (g)**, and a circular mask is applied in **Step (h)** to isolate the solar disk, removing any background data outside the Sun's circumference. Finally, **Step (i)** shows the active region (AR) and flare location (FL) points, indicating whether they fall within the defined bounding regions. This sequence illustrates the proximity-based analysis used to evaluate the alignment between the model's predictions and actual solar activity.

B. Explanation Evaluation Metrics

To quantify the explanations' relevance to AR locations, we introduce two metrics: (i) Proximity Score (PS) shown in Eq. (4), and Attribution Colocation Ratio (ACR) shown in Eq. (5). The PS quantifies the average Euclidean distance

a. GGCAM Activation Map   b. Edge Detection   c. Clustered points

d. Convex Hulls   e. East-West Buffers   f. Reclustered Buffers

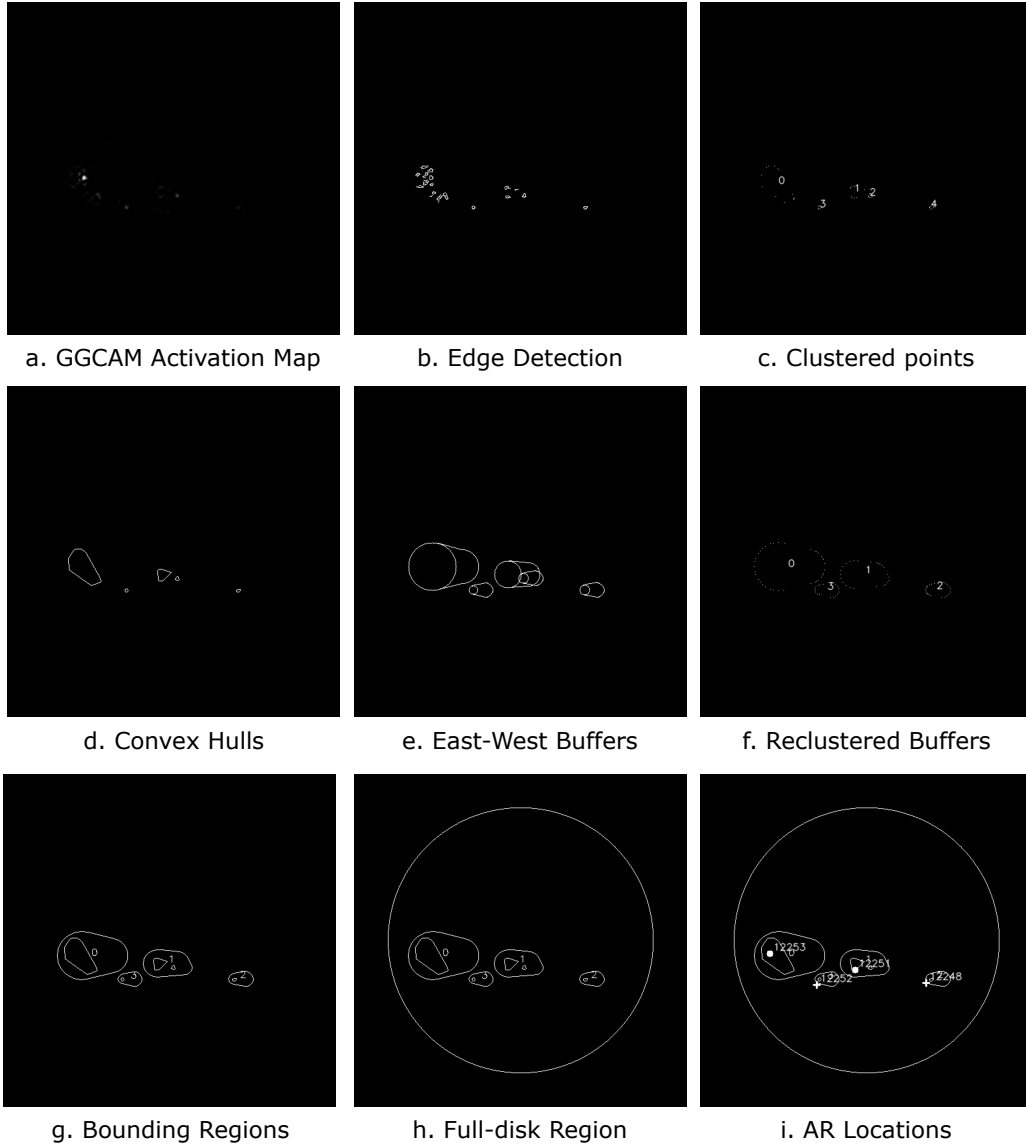g. Bounding Regions   h. Full-disk Region   i. AR Locations

Fig. 3. An example of sequential processing applied to the attribution maps of an input magnetogram captured on 2015-01-01 at 07:59:39.30 UTC.

between observed ARs and their respective closest convex hulls and ACR measures the percentage of ARs located within the bounding buffer of the total ARs present in full-disk image. To elaborate, PS measures how close the active regions are to the predicted regions, providing insight into the precision of the model's predictions. ACR, on the other hand, indicates whether the active regions are correctly located within the predicted areas. Together, PS offers a continuous measure of distance, while ACR provides a binary measure of correctness. Using both metrics allows for a more comprehensive evaluation, with PS focusing on precision and ACR ensuring that the model captures the relevant areas.

$$PS = \frac{\sum_{i=1}^{n} d_{\min}(AR_i, AM_i)}{n} \quad (4)$$

$$ACR = \frac{\sum_{i=1}^{n} \text{inside\_bounding\_box}(AR_i)}{n} \times 100 \quad (5)$$

where AR_i denotes the ith AR, AM_i represents the closest convex hull to AR_i, n is the total number of ARs, and inside_bounding_box(AR_i) is a binary function that equals 1 if AR_i is within the bounding buffer, and 0 otherwise.

III. EXPERIMENTAL EVALUATION

The focus of this study is to evaluate how well the activated pixels in post hoc explanations align with the locations of active regions (ARs) and flares, particularly X- and M-class flares, which are significant for Earth-impacting events. The proximity-based approach quantifies the quality of these explanations and assesses the models' performance. For our evaluation, the dataset comprises 5,923 full-disk line-of-sight (LoS) solar magnetogram images. These magnetograms, sampled every four hours from January 2020 to December 2023, are labeled with a 24-hour prediction window based on the maximum peak X-ray flux. As mentioned earlier, we used two

TABLE I
PARAMETERS USED FOR PROXIMITY ANALYSIS OF M1 AND M2 MODELS

Parameter	Description	Value
lower_threshold	Lower Pixel Intensity Threshold	30
upper_threshold	Upper Pixel Intensity Threshold	50
min_samples	Minimum Samples for Clustering	2
max_dist	Maximum Clustering Distance (pixels)	10
eastward_buffer	Eastward Buffer (pixels)	5
westward_buffer	Westward Buffer (pixels)	40
target_size	Target Image Size (pixels)	512
scale_to	Pixel Intensity Scale Factor	255

TABLE II
AVERAGE PS AND ACR FOR M1 AND M2 MODELS

Model	Average PS	Average ACR
M1	61.405091	0.552943
M2	29.420010	0.550404

full-disk models presented in [6], denoted as model M1, and model M2, presented in [18]. We used GGCAM-generated attribution maps to perform a proximity analysis, comparing the attribution alignment of the M1 and M2 models with the actual locations of active regions (ARs) and flares. This analysis, guided by the parameters detailed in Table I, provided critical insights into the models' performance and reliability in predicting solar flares.

While the predictive performance of models M1 [6] and M2 [18] is reported in the respective studies, we present an evaluation of the explanations in terms of the average Proximity Score (PS) and Attribution Colocation Ratio (ACR) generated using these models, as shown in Table II. We observed that M2 demonstrates better spatial alignment with actual flare locations, as indicated by a lower PS, while both models perform similarly in terms of ACR. Fig. 4 illustrates the distribution of proximity scores across different classification categories (TN, FP, TP, FN) for M1 and M2 respectively. The results indicate that M2 generally produces lower median proximity scores with less variability, suggesting more consistent and accurate predictions of flare regions compared to M1. The source code and other necessary artifacts are available from our open-source repository [20].

IV. DISCUSSION

A. Comparative Analysis of PS Across Models

TABLE III
SUMMARY OF PROXIMITY-BASED SCORES FOR M1 AND M2 MODELS

Metric	Mean		Standard Deviation	
	M1	M2	M1	M2
FN	72.49	38.02	71.89	46.96
FP	52.95	55.63	52.72	49.05
TN	61.70	28.53	66.59	40.26
TP	59.43	52.03	68.05	48.68

The comparative analysis of proximity-based scores for the M1 and M2 models (Table III) reveals significant differences in their explanation alignment with the ground truth. M1 generally shows higher mean proximity scores and greater
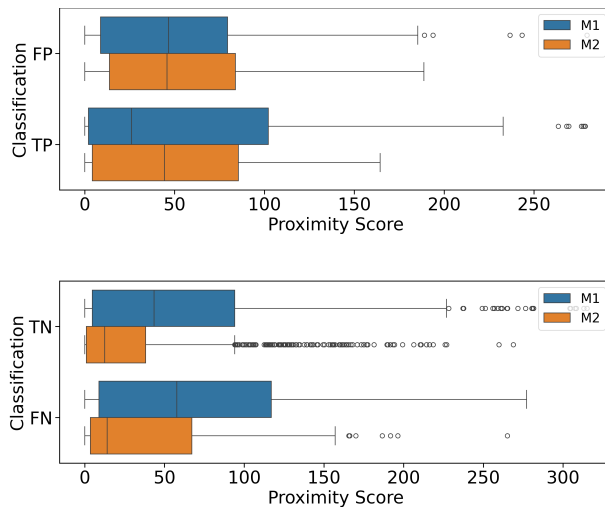

Fig. 4. Distribution of Proximity Scores, grouped in terms of four outcomes of contingency matrix (referred in text as categories). (a) Proximity score distribution for True Positives (TP) and False Positives (FP) (b) Proximity score distribution for False Negatives (FN) and True Negatives (TN).
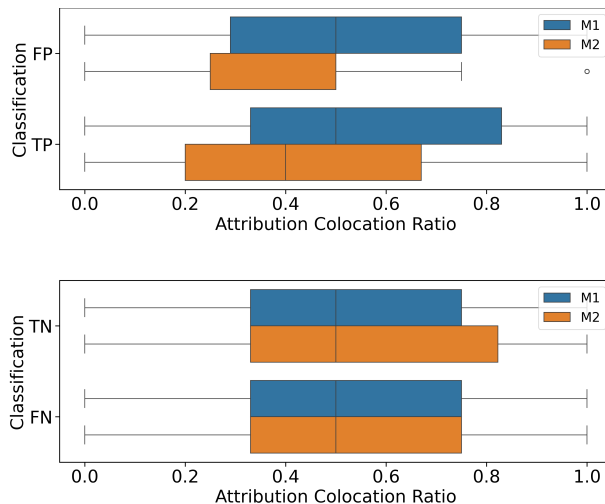

Fig. 5. Distribution of Attribution Colocation Ratios for both models, grouped in terms of four outcomes of contingency matrix (referred in text as categories). (a) ACR distribution for True Positives (TP) and False Positives (FP) (b) ACR distribution for False Negatives (FN) and True Negatives (TN).

variability across all categories, indicating that its predictions are spatially farther from the true flare regions. In contrast, M2 demonstrates lower mean scores with reduced variability, reflecting more accurate and consistent predictions. Specifically, M2 outperforms M1 in false negatives, false positives, and true negatives, highlighting its improved alignment with actual flare regions. True positive scores for M2 further underscore its better spatial accuracy, making it more reliable for solar flare prediction. The box plots (Fig. 4) illustrate the distribution of proximity scores across different categories, offering further insights. M1 displays higher median scores and greater variability, suggesting less precise predictions. Conversely, M2 exhibits lower medians and fewer extreme outliers, indicating closer and more consistent predictions. For true negatives, M2 achieves significantly lower median

proximity scores, showcasing superior alignment with actual flare regions. Both models perform similarly for false positives and false negatives, but M2's reduced variability and fewer outliers underscore its consistency.

B. Comparative Analysis of ACR Across Models

TABLE IV
SUMMARY OF ATTRIBUTION COLOCATION RATIOS FOR M1 AND M2

Metric	Mean		Standard Deviation	
	M1	M2	M1	M2
FN	0.514	0.521	0.311	0.310
FP	0.511	0.446	0.289	0.273
TN	0.561	0.562	0.305	0.328
TP	0.574	0.463	0.306	0.342

The ACR analysis for the M1 and M2 models (Table IV) highlights notable differences in prediction accuracy and consistency. M1 generally exhibits higher mean ACRs across most categories, suggesting that its predictions are more closely aligned with actual flare regions. However, M2 demonstrates less variability in these ratios, indicating a more consistent alignment across its predictions. The comparative mean and standard deviation values provide further insight into each model's relative performance and consistency in predicting solar flare locations. The box plots (Fig. 5) offer additional detail on the distribution of ACRs. While both models display similar medians across all categories, M2 shows narrower interquartile ranges (IQRs) for false positives and true negatives, reflecting more consistent alignment. In contrast, M1 displays greater variability in true positives and false negatives, suggesting less consistent predictions in these categories. For true positives, M2 achieves a lower median ACR but with reduced variability, implying consistent but slightly less accurate predictions compared to M1. For false negatives, M1 shows a slightly higher median but greater variability, further underscoring its inconsistency. Overall, M1 has higher median ACRs in some categories but less consistency due to greater variability, while M2 is more consistent with reduced variability.

V. CONCLUSION

This study introduces a novel and fully automated method for analyzing explanations from full-disk models in a near-real-time setting. This approach represents a significant step forward in seamlessly integrating explainability into operational systems, enhancing the transparency and reliability of deep learning models for solar flare prediction. We utilized this method to evaluate the explanations generated from two CNN-based models and compared the reliability of these models. The findings underscore the potential of these explainable methods to bridge the gap between advanced model outputs and actionable insights. Future work should prioritize refining post-hoc explanation techniques, with an emphasis on mitigating noise during preprocessing, to further improve model interpretability and operational utility.

REFERENCES

[1] 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*, vol. 159, no. 1, pp. 19–106, Aug. 2011.

[2] Y. Yasyukevich, E. Astafyeva, A. Padokhin, V. Ivanova, S. Syrovatskii, and A. Podlesnyi, "The 6 september 2017 x-class solar flares and their impacts on the ionosphere, gnss, and hf radio wave propagation," *Space Weather*, vol. 16, Jul. 2018.

[3] S. Toriumi and H. Wang, "Flare-productive active regions," *Living Reviews in Solar Physics*, vol. 16, no. 1, p. 3, May 2019.

[4] C. Pandey, R. Angryk, and B. Aydin, "Solar flare forecasting with deep neural networks using compressed full-disk hmi magnetograms," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 1725–1730.

[5] C. Pandey, A. Ji, R. A. Angryk, M. K. Georgoulis, and B. Aydin, "Towards coupling full-disk and active region-based flare prediction for operational space weather forecasting," *Frontiers in Astronomy and Space Sciences*, vol. 9, Aug. 2022.

[6] C. Pandey, A. Ji, T. Nandakumar, R. A. Angryk, and B. Aydin, "Exploring deep learning for full-disk solar flare prediction with empirical insights from guided grad-cam explanations," 2023.

[7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

[8] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014.

[9] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," 2017.

[10] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.

[11] S. Bhattacharjee, R. Alshehhi, D. B. Dhuri, and S. M. Hanasoge, "Supervised convolutional neural networks for classification of flaring and nonflaring active regions using line-of-sight magnetograms," *Astrophysical Journal*, vol. 898, no. 2, Aug. 2020.

[12] C. Pandey, R. A. Angryk, and B. Aydin, *Explaining full-disk deep learning model for solar flare prediction using attribution methods*. Springer Nature Switzerland, 2023, pp. 72–89.

[13] C. Pandey, A. Ji, R. A. Angryk, and B. Aydin, "Towards interpretable solar flare prediction with attention-based deep neural networks," in *2023 IEEE Sixth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2023, pp. 83–90.

[14] Z. S. et al., "Predicting solar flares using cnn and lstm on two solar cycles of active region data," *The Astrophysical Journal*, vol. 931, no. 2, p. 163, 2022.

[15] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," 2017.

[16] C. Pandey, R. A. Angryk, M. K. Georgoulis, and B. Aydin, *Explainable deep learning-based solar flare prediction with post hoc attention for operational forecasting*. Springer Nature Switzerland, 2023, pp. 567–581.

[17] C. Pandey, R. A. Angryk, and B. Aydin, *Deep neural networks based solar flare prediction using compressed full-disk line-of-sight magnetograms*. Springer International Publishing, 2022, pp. 380–396.

[18] C. Pandey, R. Angryk, and B. Aydin, "Unveiling the potential of deep learning models for solar flare prediction in near-limb regions," in *2023 International Conference on Machine Learning and Applications (ICMLA)*, 2023, pp. 703–708.

[19] K. Yi, Y.-J. Moon, D. Lim, E. Park, and H. Lee, "Visual explanation of a deep learning solar flare forecast model and its relationship to physical parameters," *The Astrophysical Journal*, vol. 910, no. 1, p. 8, 2021.

[20] DMLAB, "Attribution-based proximity analysis," 2024. [Online]. Available: https://bitbucket.org/gsudmlab/solar-flare-prediction-evaluation/src/main/