Integrating Full-Disk and Active-Region Models for Reliable Forecasting of Solar Flares

Duc Le, Robin Chataut, Chetraj Pandey

Texas Christian University

Introduction

- Solar flares are major drivers of space-weather disturbances and can disrupt satellites, communication systems, and navigation infrastructure.
- Deep learning methods have shown promising capabilities in flare forecasting, yet many models operate either on full-disk solar observations or on isolated active-region patches. This separation limits their ability to combine global solar context with localized magnetic structure and affect the reliability of predictions. In addition, full-disk models often provided limited information about which regions drive their forecasts.
- This projects develops a two-stage deep learning pipeline that integrates full-disk and active-region level analysis within a unified flare forecasting pipeline.

Models and Data

- In this study, we use the deep learning-based solar flare prediction model [1] trained with standardized images [1, 4] from HMI SHARP data series. It takes a full-disk magnetogram as input and outputs a global flare probability (Fig. 1) .

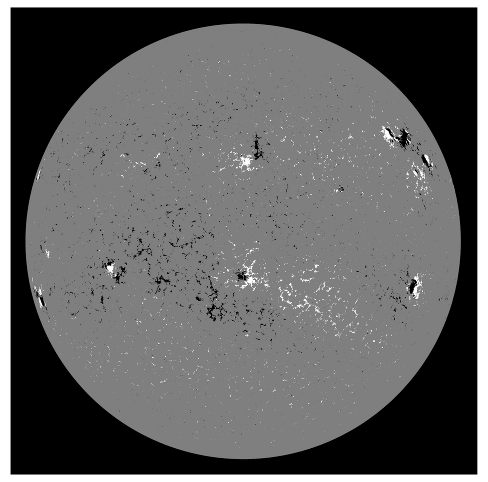

Fig. 1. Full-disk solar magnetogram shown in grayscale and resized to 512 × 512 for input to the full-disk model.

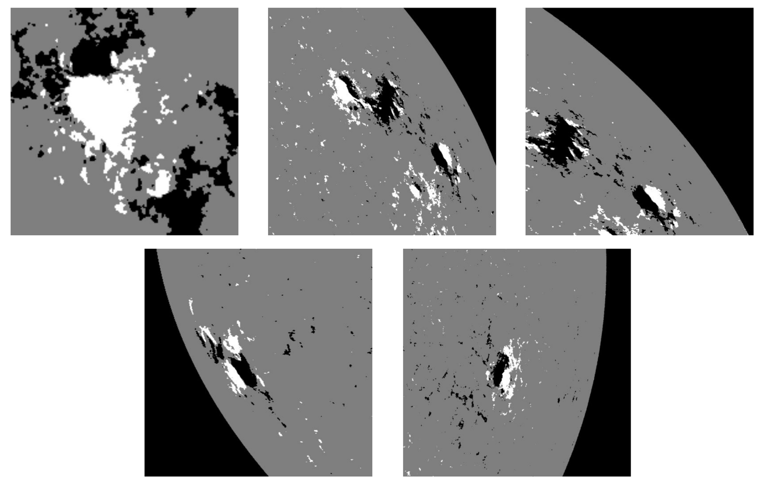

Fig. 2. Top candidate regions identified by Guided Grad-CAM, Integrated Gradients, and DeepLiftShap, with final significance validation using Occlusion.

- We next apply attribution methods to identify the most influential regions driving the global full-disk prediction. These highlighted regions are projected back onto the original solar magnetogram, consolidated into candidate areas, and cropped for follow-up analysis (Fig. 2). The resulting patches are then passed to a trained ResNet-based active-region mode [5]. Each patch is converted to a 512 × 512 grayscale image, and the model produces a localized flare probability for each candidate region.

Methodology: Full-Disk to Active-Region Prediction Pipeline

- A full-disk solar magnetogram in FITS format is first preprocessed into a grayscale 512 × 512 JPG image for input to the full-disk model. As shown in Fig. 3, this preprocessing step preserves the large-scale magnetic structure of the solar disk while removing non-solar background pixels. The full-disk model then produces a global flare probability, which serves as the first stage of the prediction pipeline.
- We then generate attribution maps that highlight the regions that most influenced the global prediction. Regions that appear consistently across methods are treated as stronger candidates and further validated using Occlusion (Fig. 4).
- The selected regions are then mapped back to the original FITS image, extracted as active-region patches, and passed to the active-region model for localized prediction.

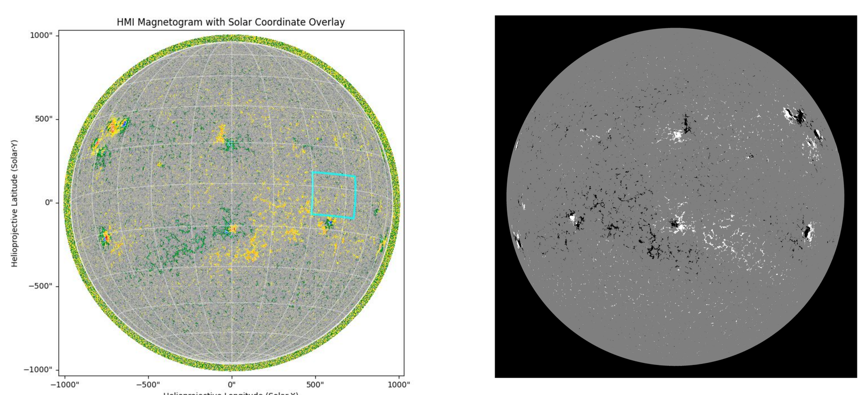

Fig. 3. Preprocessing of a full-disk solar magnetogram for full-disk model input. The left panel shows the original HMI magnetogram in FITS format with solar coordinate context. The right panel shows the corresponding preprocessed grayscale 512 x 512 JPG image used as input to the full-disk model. In the processed image, white indicates strong positive magnetic field values, black indicates strong negative values, gray indicates near-neutral regions, and pixels outside the solar disk are masked and shown in black.

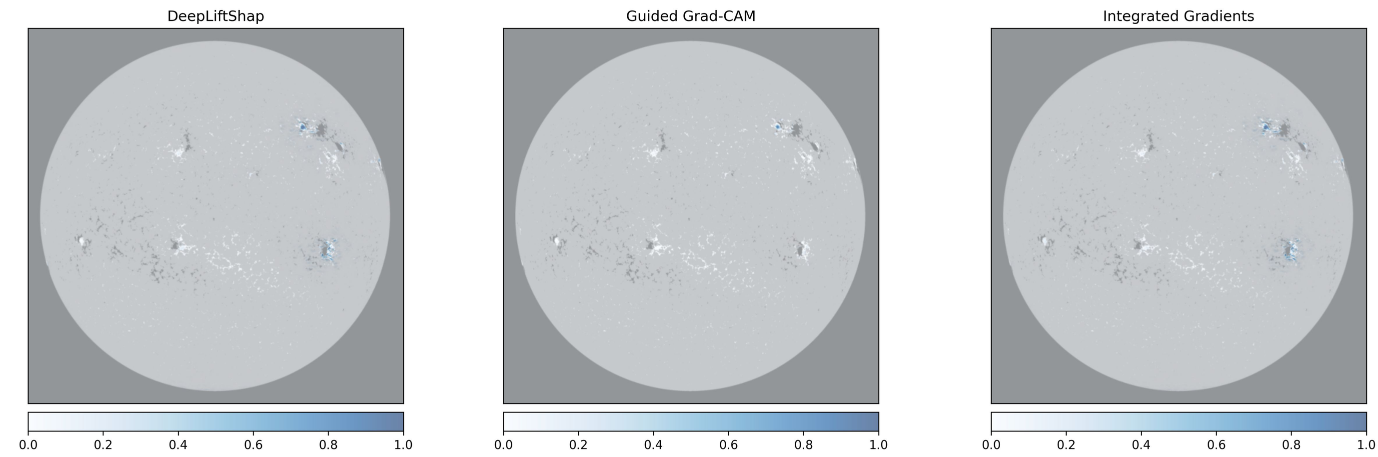

Fig. 4. Attribution maps generated by DeepLiftShap, Guided Grad-Cam, and Integrated Gradients for a full-disk solar magnetogram. Each method highlights regions that contributed the most to the global flare prediction. Brighter highlighted areas indicate stronger influence on the model output. The overlap across the three methods suggests consistent hotspots regions, which are then used as candidate locations for further active-region analysis.

System Interface and Architecture

- The Streamlit dashboard (Fig. 5) displays the full-disk magnetogram, global flare probability, attribution heatmaps, active-region patches, localized patch predictions, and past prediction history.
- To support continuous forecasting, the prediction pipeline is designed to run automatically every hour through Amazon Evenbridge, which trigger a containerized prediction job, as shown in (Fig. 6).

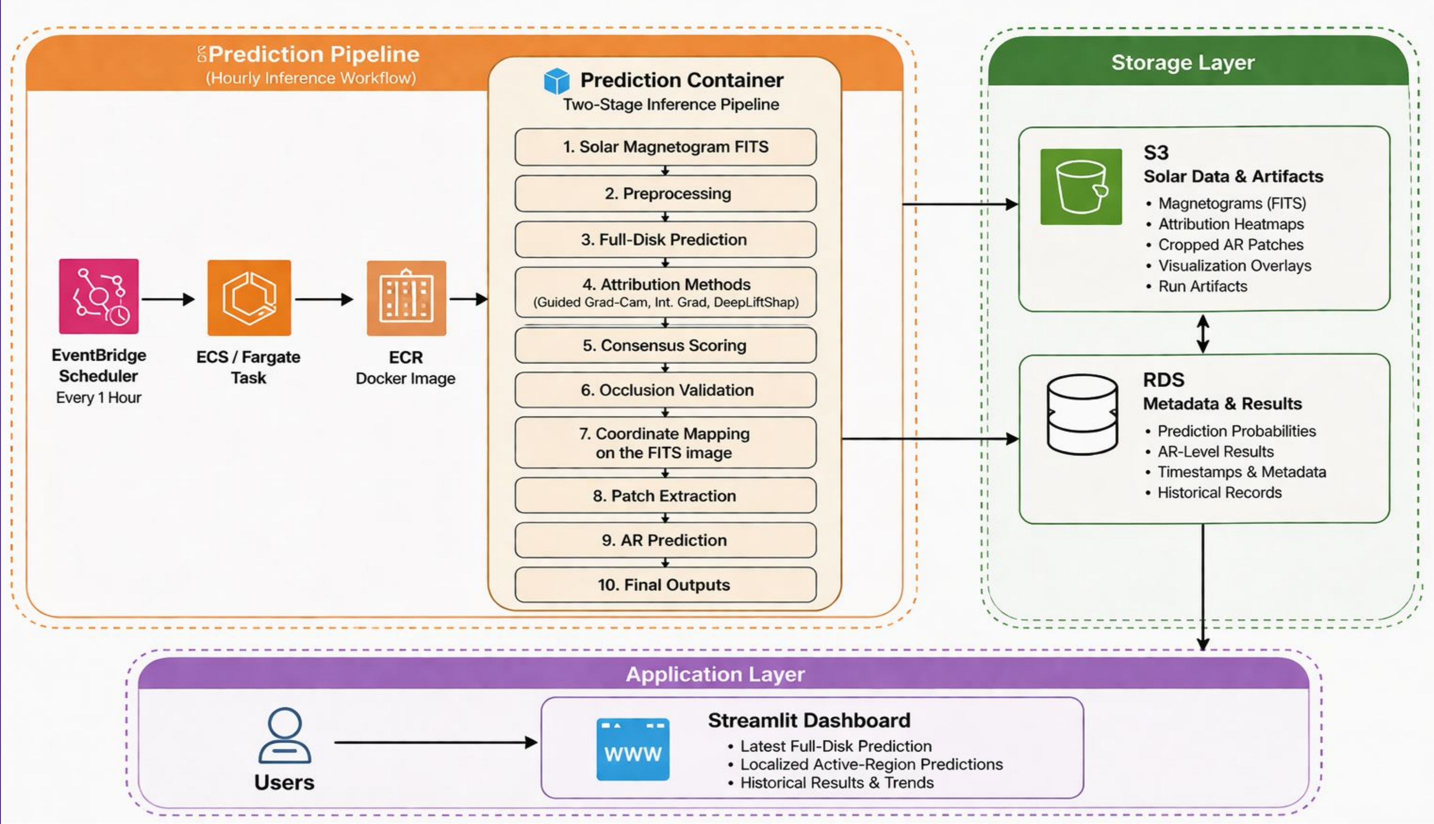

Fig. 6. Architecture of the 2-stage solar flare prediction system. An hourly EventBridge scheduler triggers an ECS/Fargate task that runs the prediction container from ECR. The container executes the pipeline in order: FITS magnetogram input, preprocessing, full-disk prediction, attribution, consensus scoring, occlusion validation, coordinate mapping, patch extraction, and AR prediction. Generated artifacts such as magnetograms, heatmaps, overlays, and cropped patches are stored in S3, while prediction results and metadata are stored in RDS. A Streamlit dashboard reads from S3 and RDS to display the latest full-disk prediction, localized active-region results, and prediction history.

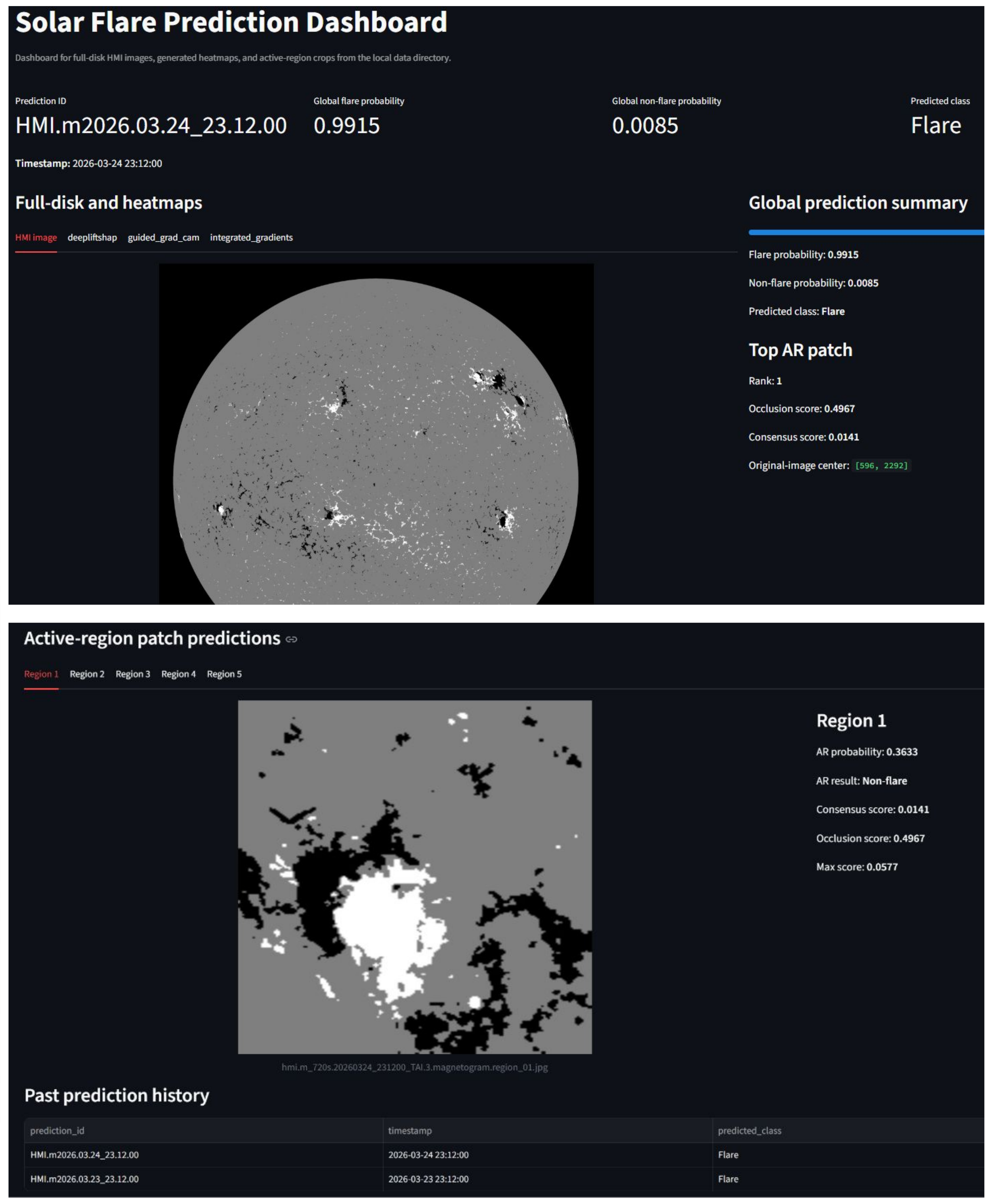

Fig. 5. Streamlit dashboard for visualizing full-disk and localized solar flare predictions. The dashboard displays the input HMI magnetogram, global flare and non-flare probabilities, the predicted class, and attribution heatmaps. It also summarizes the top-ranked active-region patch selected for follow-up localized predictions., including its rank, occlusion score, consensus score, and original-image location.

Remarks and Future Directions

- The results suggest that this framework can produce more informative and interpretable flare forecasts than a global-only model by combining full-disk prediction with localized active-region analysis.
- However, the quality of the localized stage depends on accurate hotspot identification from the attribution maps.
- **Future direction:** Future work will focus on improving candidate-region selection, validating robustness across diverse events, and strengthening the deployment pipeline for reliable real-time forecasting.

References

1. C. Pandey, T. Adeyeha, J. Hong, R. A. Angryk, and B. Aydin, "Advancing solar flare prediction using deep learning with active region patches," in Machine Learning and Knowledge Discovery in Databases: ADS Track, Springer Nature Switzerland, 2024, pp. 50–65, doi: 10.1007/978-3-031-70381-2_4.
2. C. Pandey, A. Ji, J. Hong, R. A. Angryk and B. Aydin, "Embedding Ordinality to Binary Loss Function for Improving Solar Flare Forecasting," 2024 IEEE 11th International Conference on Data Science and Advanced Analytics (DSAA), San Diego, CA, USA, 2024, pp. 1-10, doi: 10.1109/DSAA61799.2024.10722839.
3. 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," in Discovery Science, Springer Nature Switzerland, 2023, pp. 567–581, doi: 10.1007/978-3-031-45275-8_38.
4. C. Pandey and B. Aydin, "Bitmap filtered line of sight HMI active region patches with augmentations," Harvard Dataverse, 2024. doi: 10.7910/DVN/WLTBQQ
5. C. Pandey, T. Adeyeha, J. Hong, R. A. Angryk, and B. Aydin, "Advancing Solar Flare Prediction Using Deep Learning with Active Region Patches," *Lecture Notes in Computer Science*, pp. 50–65, doi: 10.1007/978-3-031-70381-2_4.

Acknowledgements

This work is supported by the Department of Computer Science at Texas Christian University, Fort Worth, Texas. The data used in this project is a courtesy of NASA/SDO and the AIA,EVE, and HMI science teams, and the NOAA (NGDC).

Contact Information

Duc Le
Undergraduate Student
B.S. Computer Science & B.A. Mathematics
Davis College of Science and Engineering
Texas Christian University
Email: duc.anh.le@tcu.edu
Website: https://github.com/DucLe-2005