CrossMark

An End-to-end Ensemble Machine Learning Approach for Predicting High-impact Solar Energetic Particle Events Using Multimodal Data

Pouya Hosseinzadeh , Soukaina Filali Boubrahimi , and Shah Muhammad Hamdi
Department of Computer Science, Utah State University, Logan, UT, USA; pouya.hosseinzadeh@usu.edu, soukaina.boubrahimi@usu.edu, s.hamdi@usu.edu

Abstract

Solar energetic particle (SEP) events, in particular high-energy-range SEP events, pose significant risks to space missions, astronauts, and technological infrastructure. Accurate prediction of these high-impact events is crucial for mitigating potential hazards. In this study, we present an end-to-end ensemble machine learning (ML) framework for the prediction of high-impact ∼100 MeV SEP events. Our approach leverages diverse data modalities sourced from the Solar and Heliospheric Observatory and the Geostationary Operational Environmental Satellite integrating extracted active region polygons from solar extreme ultraviolet (EUV) imagery, time-series proton flux measurements, sunspot activity data, and detailed active region characteristics. To quantify the predictive contribution of each data modality (e.g., EUV or time series), we independently evaluate them using a range of ML models to assess their performance in forecasting SEP events. Finally, to enhance the SEP predictive performance, we train an ensemble learning model that combines all the models trained on individual data modalities, leveraging the strengths of each data modality. Our proposed ensemble approach shows promising performance, achieving a recall of 0.80 and 0.75 in balanced and imbalanced settings, respectively, underscoring the effectiveness of multimodal data integration for robust SEP event prediction and enhanced forecasting capabilities.

1. Introduction

Solar energetic particle (SEP) events are a significant challenge in space-weather forecasting, due to the considerable risks they pose to space missions, astronauts, and technological infrastructure (S. Kahler 1982; D. V. Reames 2013). High-energy particles are generated during intense solar events, such as solar flares and coronal mass ejections (CMEs; N. Gopalswamy et al. 2002), and they can travel rapidly through the heliosphere, creating dangerous conditions both in space and on Earth. Accurate predictions of SEP events is essential, as SEPs can disrupt communication systems, damage satellite electronics, and expose astronauts to harmful radiation, especially during extravehicular activities (EVAs; D. Smart & M. Shea 2003; S. Furukawa et al. 2020). SEP events also pose a risk to electronic systems on Earth, as they can interfere with power grid operations and navigation systems, which we rely on daily for transportation, communication, and emergency response (C. J. Schrijver & G. L. Siscoe 2010). Additionally, the consequences of SEP events are not limited to Earth; they are also a major concern for future lunar and Martian missions. On the Moon, the lack of an atmosphere and a global magnetic field means that astronauts are directly exposed to solar radiation during surface activities, making SEP prediction critical to ensure their safety during EVAs. Similarly, Mars has only a weak magnetic field and a thin atmosphere, which offer limited protection against SEPs. As a result, both astronauts and robotic explorers on Mars are vulnerable to radiation hazards, emphasizing the necessity for reliable SEP prediction models to support long-term human exploration beyond Earth.

SEP events are composed of a mix of particles, including protons, electrons, and heavy ions, which are accelerated to high energies by solar flares and CMEs (I. Richardson & H. Cane 2010). These particles can reach Earth within minutes to hours, depending on their energy levels and the intensity of the parent solar event (i.e., solar flare and/or CME). Additionally, SEP events exhibit a wide range of energy levels, spanning from keV to GeV. Within this spectrum, high-impact SEP events emerge as particularly significant due to their capacity to induce substantial disruptions to space missions, astronauts, and technological infrastructure. One particularly intense manifestation of an SEP event is a solar proton event (SPE), which involves a substantial increase in the flux of high-energy protons detected by space-based instruments. SPEs are often identified when the flux of >10 MeV protons exceeds 10 pfu, a threshold established to classify such events operationally as S1 events. When SPEs reach relativistic energies, they can penetrate Earth's atmosphere, producing secondary particles that significantly increase the count rates of ground-based detectors, an occurrence known as a ground-level enhancement (S. E. Forbush 1946).

The study of solar activity and its impacts on space weather has a rich history, dating back to the work of astronomers such as Schwabe, Sabine, and Carrington in the 19th century (R. C. Carrington 1859). Carrington's meticulous observations in the mid-1800s not only revealed key aspects of solar dynamics, including the Sun's rotation axis and latitudinal variation of sunspots over the solar cycle (D. H. Hathaway 2015), but also provided crucial insights into the connection between solar phenomena and geomagnetic disturbances (E. W. Cliver & L. Svalgaard 2004). On 1859 September 1, Carrington made the first visual observation of a solar flare, marking a significant milestone in the study of solar activity. This event, later termed the "Carrington event," was characterized by intense white light patches erupting from the Sun's surface, followed by marked variations in geomagnetic activity (B. T. Tsurutani et al. 2003). While the causal relationship between solar eruptions and geomagnetic disturbances was not immediately established, subsequent research, notably by

Bartels in the 1930s, elucidated the mechanisms underlying such phenomena (E. E. Davies et al. 2021). Over the decades, advancements in observational techniques and theoretical understanding have deepened our knowledge of SEP events. Notably, early observations, such as ground-based measurements of ionization chambers and neutron monitors, laid the groundwork for recognizing the association between solar flares and energetic particles observed at Earth (S. E. Forbush 1946). However, it was not until the mid-20th century that a clearer understanding of SEP event categorization emerged, distinguishing between impulsive and gradual events based on their observational characteristics (D. V. Reames 1999). This two-class paradigm provided a framework for interpreting observations of SEP events at various stages of their evolution, from their origins near the Sun to their arrival at Earth (N. Gopalswamy et al. 2014). In recent years, significant advancements in our ability to observe and understand SEP events have been achieved through multispacecraft observations and theoretical modeling efforts (J. Mazur et al. 2002; I. Richardson et al. 2014; M. Desai et al. 2016). For instance, observations from the Solar and Heliospheric Observatory (SOHO) and the Advanced Composition Explorer have provided data on particle charge states and elemental abundances, allowing researchers to trace SEP particles back to their sources and understand their acceleration mechanisms (R. Mewaldt 2013). On the modeling side, advancements in particle transport simulations, such as those developed by G.-M. Le & X.-F. Zhang (2017), have improved our understanding of how SEP particles propagate through complex magnetic fields in the heliosphere.

There are two different noncompeting paradigms for solar events forecasting: physics-based models of heliophysics processes and machine learning (ML) models. Physics-based models have been fundamental in the early development of SEP event prediction. Physics-based models are rooted in our understanding of solar and space physics principles, incorporating plasma physics, magnetohydrodynamics (MHD), and particle transport theories to simulate and predict the behavior of SEP events. For instance, SOLPENCO predicts the flux and fluence of gradual SEP events based on CME characteristics by applying MHD principles to model CME-driven shock acceleration and transport dynamics (A. Aran et al. 2006). Similarly, the physics-based model by T. Sato et al. (2018) leverages real-time observations to provide warnings for SEP events, employing kinetic theory and particle propagation equations to forecast SEP fluxes. While physics-based models offer valuable insights and predictions, they are often limited by computational demands and the need for real-time observational data, which can restrict their operational use. On the other hand, ML models are trained on historical observations to identify patterns and relationships within the data, enabling data-driven insights and predictions (E. Camporeale 2019). ML models can process vast amounts of historical data to identify precursors of SEP events and forecast future occurrences. Boubrahimi et al. employed a decision tree ML model to provide short-term warnings of > 100 MeV SEP events using X-ray and proton data, resulting in a reduced false-alarm rate (S. F. Boubrahimi et al. 2017). Similarly, Stumpo et al. and Laurenza et al. optimized logistic regression models with input parameters like flare heliolongitude and synthetic minority oversampling to enhance SEP prediction accuracy (M. Laurenza et al. 2009; M. Stumpo et al. 2021). More recently, there have been several studies based on time-series (TS) data where an observation window is set before the occurrence of a solar flare associated with an SEP event (P. Hosseinzadeh et al. 2023, 2024a, 2024b). For instance, Boubrahimi et al. showed that multivariate TS of X-ray and proton flux data could effectively serve as precursors for >100 MeV SEP events (S. F. Boubrahimi et al. 2017). Aminalragia-Giamini et al. trained a deep learning model on soft X-ray measurements of solar flares to predict SEP onsets, achieving notable accuracy (S. Aminalragia-Giamini et al. 2021). The UMASEP model also incorporates ML techniques in real-time SEP predictions, operating at the CCMC (M. Núñez & D. Paul-Pena 2020). Lavasa et al. further assessed multiple ML models, such as logistic regression, MLP neural networks, random forests, SVMs, decision trees, and extreme gradient boosting, for SEP forecasting (E. Lavasa et al. 2021). Additionally, Stumpo et al. applied SVM and linear regression methods, achieving performance comparable to the ESPERTA model (M. Stumpo et al. 2021). Recent work has demonstrated that data augmentation techniques can enhance SEP predictions, offering improved performance in ML models (P. Hosseinzadeh et al. 2023, 2024a, 2024b, 2024c; S. Filali Boubrahimi et al. 2024; P. Hosseinzadeh et al. 2024c; S. Kasapis et al. 2024; P. A. Kosovich et al. 2024). Over the past year, there has been a growing body of research exploring ML-based approaches for SEP prediction, reflecting the increasing interest in leveraging data-driven methods to enhance forecasting accuracy and operational readiness (S. Kasapis et al. 2022; M. Nedal et al. 2023; A. Ali et al. 2024; S. Chatterjee et al. 2024; M. Laurenza et al. 2024; I. Jackson & P. Martens 2024; S. Rotti & P. C. Martens 2024; S. A. Rotti et al. 2024).

In this study, we target the integration of diverse data modalities to improve the prediction of high-energy SEP events. By leveraging solar X-ray imagery, proton flux measurements, sunspot activity, and active region characteristics sourced from the Geostationary Operational Environmental Satellite (GOES) and SOHO, this study addresses the complexities of integrating multimodal data for a more robust prediction framework. This comprehensive approach enhances both the accuracy and the time window of early-warning systems critical for space-weather forecasting. Our contributions are as follows:

1. Developing a multimodal data integration framework that combines solar EUV images, proton flux data, sunspot activity, and active region features to capture both spatial and temporal dynamics relevant to SEP events.
2. Implementing multiple ensemble ML models that optimally combines information from each data modality, improving predictive performance and reliability over single-modality approaches.
3. Enhancing early-warning capabilities by leveraging diverse data sources, allowing for more accurate predictions of high-energy SEP events, thus supporting proactive space-weather forecasting and risk mitigation efforts.

The rest of this paper is organized as follows: Section 2 introduces the data set used in this study, detailing the integration of diverse data modalities, such as tabular data, EUV active region polygons, and GOES TS proton flux, which collectively capture the critical aspects of solar activity pertinent to SEP event prediction. Section 3 outlines the methodology, describing the various ML models and ensemble learning strategies employed. Section 4 presents experimental

Table 1
Summary of Data Modalities Used in This Study

Data Modality	Description	Data Source	Preprocessing
Tabular Data	Solar event information, including sunspot counts, active region (AR) counts, and flare class, capturing trends in solar activity before SEP events	HEK (LMSAL)	None
EUV Active Region Polygon Data	Spatial details of ARs from EUV images, with polygon shapes and AR boundaries, highlighting regions of activity prior to SEP events	Helioviewer (SOHO EIT 304 Å)[a]	Gray-scale conversion, edge detection, binary masking
GOES Time-series Proton Flux Data	Proton flux levels over time from the GOES satellite, specifically in the P6 channel (84-200 MeV), showing flux variations associated with SEP events	NOAA NCEI[b]	Z-score normalization

Note.
[a] https://helioviewer.org/
[b] https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/

results, analyzing the effectiveness of each model on different data modalities and identifying optimal configurations for ensemble models. Finally, Sections 5 and 6 provide a comprehensive discussion of the findings, conclude the paper, and suggest future research directions for this work.

2. Data Set

In our study, we integrate diverse data modalities, each offering valuable insights to advance the prediction of high-energy SEP events. Specifically, we utilized three data modalities: tabular data capturing solar event characteristics, spatial information derived from extreme ultraviolet (EUV) imaging of active regions, and TS proton flux data from GOES. By combining these three data modalities, we aim to capture both the temporal and spatial dynamics of solar activity, which are crucial for identifying patterns that precede high-energy-range SEP events. Table 1 provides a summary of each data modality, including descriptions, sources, and preprocessing techniques applied to enhance data quality for the analysis. The following sections outline the process for determining SEP and non-SEP events, followed by a detailed description of each data modality, the data collection process, and the preprocessing pipeline.

2.1. SEP and Non-SEP labels

Similarly to previous studies (M. Núñez 2011; S. F. Boubrahimi et al. 2017; P. Hosseinzadeh et al. 2024b), in this work, we define non-SEP events as X-ray flares with a peak intensity of at least C1.3 that did not result in any SEP event. This classification assumes that a ~100 MeV impulsive event may occur if the corresponding X-ray flare reaches a peak intensity of C1.3 or higher. For example, an X-class flare on 2023 May 12 with no detected proton activity or an M5.6 flare on 2022 November 3 without SEP association would both be categorized as non-SEP events. We collected SEP events from the GSEP catalog (S. Rotti et al. 2022) and sourced non-SEP events from the Heliophysics Event Knowledgebase (HEK).[1] We used 37 SEP events and collected another 37 non-SEP events to create a balanced data set. The slightly smaller number of SEP events in our study, compared to other research, is primarily due to the unavailability or inconsistency of multimodal data for certain events. Our approach requires complete data from all modalities for each event, and if any modality is missing or

incomplete for a specific SEP event, that event is excluded. Note that non-SEP events were randomly selected to ensure a fair and unbiased comparison with the SEP events. To maintain consistency, we ensured that the non-SEP events spanned the same years as the SEP events. Additionally, we excluded non-SEP events that occurred too close in time to SEP events, to avoid potential overlap in input data. The full list of SEP events analyzed in this study is shown in Table 2 for reference. Also, our Python source code is publicly accessible on GitHub.[2]

2.2. Tabular Data

We collected tabular data under the form of a vector of background solar conditions during the time of the solar flare event. These features were obtained from the HEK, also called LMSAL, which compiles detailed data on solar events across multiple solar cycles. This data set captures solar activity leading up to and surrounding solar flare events. The tabular data includes sunspot counts, active region (AR) counts, and solar flare classes. The sunspot numbers serve as a key indicator of solar magnetic activity. Sunspot numbers refer to the total count of individual sunspots visible on the solar disk, while AR counts refer to the number of distinct magnetically active regions, each of which may contain multiple sunspots. This distinction is important, as a single active region can host multiple sunspots, leading to different values for sunspot numbers and AR counts. Increased sunspot activity is often associated with heightened solar flare occurrences, and by extension, SEP events. In this data set, we count the number of sunspots occurring within the 6 hr leading up to each solar flare's start time. This approach provides a temporal context, allowing us to identify solar activity patterns that may precede high-energy SEP events. On the other hand, the AR counts indicate the number of magnetically active regions observed during the 6 hr before the onset of each solar flare. By examining AR counts in the preflare timeframe, we gain insight into heightened magnetic activity directly preceding potential high-energy SEP events. Finally, solar flare intensities, based on the GOES X-ray flare scale, serve as an indicator of the flare's energy release. The flare intensities are categorized into C-class (small), M-class (medium), and X-class (large) levels based on the GOES X-ray flare scale, which is measured on a logarithmic scale. Each class represents a tenfold increase in X-ray flux, with C-class flares indicating lower energy release, M-class flares indicating medium energy release, and

---

[1] https://www.lmsal.com/isolsearch

[2] https://github.com/pouyahosseinzadeh/Solar-Energetic-Particle-Event-Prediction-Ensemble-TS-HOG-TB

Table 2
SEP Events Used in This Study

ID	Flare Start Time	Flare Peak Time	SEP Start Time	Sunspot #	AR #	Flare Class
			SEP Event List			
1	1997-11-04 05:52:00	1997-11-04 05:58:00	1997-11-04 06:05:00	28	15	X2.1
2	1997-11-06 11:49:00	1997-11-06 11:55:00	1997-11-06 12:20:00	23	7	X9.4
3	1998-04-20 09:38:00	1998-04-20 10:21:00	1998-04-20 11:00:00	8	6	M1.4
4	1998-05-06 07:58:00	1998-05-06 08:09:00	1998-05-06 08:15:00	26	9	X2.7
5	1998-11-14 05:00:00	1998-11-14 05:08:00	1998-11-14 06:10:00	19	20	C1.3
6	2000-06-10 16:40:00	2000-06-10 17:02:00	2000-06-10 17:00:00	33	23	M5.2
7	2000-07-14 10:03:00	2000-07-14 10:24:00	2000-07-14 10:15:00	65	23	X5.7
8	2000-07-22 11:17:00	2000-07-22 11:34:00	2000-07-22 11:50:00	120	38	M3.7
9	2000-09-12 11:31:00	2000-09-12 12:13:00	2000-09-12 14:00:00	3	13	M1.0
10	2000-10-16 06:40:00	2000-10-16 07:28:00	2000-10-16 08:00:00	22	17	M2.5
11	2000-11-24 04:55:00	2000-11-24 05:02:00	2000-11-24 06:00:00	28	21	X2.0
12	2000-11-24 14:51:00	2000-11-24 15:13:00	2000-11-24 15:55:00	22	12	X2.3
13	2000-11-25 00:59:00	2000-11-25 01:31:00	2000-11-26 00:05:00	0	20	M8.2
14	2001-01-28 15:40:00	2001-01-28 16:00:00	2001-01-28 16:45:00	23	14	M1.5
15	2001-04-09 15:20:00	2001-04-09 15:34:00	2001-04-09 16:05:00	33	30	M7.9
16	2001-04-15 13:19:00	2001-04-15 13:50:00	2001-04-15 13:50:00	38	19	X14.4
17	2001-04-18 02:11:00	2001-04-18 02:14:00	2001-04-18 02:35:00	7	25	C2.2
18	2001-05-20 06:00:00	2001-05-20 06:03:00	2001-05-20 06:55:00	0	16	M6.4
19	2001-06-15 16:15:00	…	2001-06-15 15:55:00	53	31	C2.2
20	2001-09-24 09:32:00	2001-09-24 10:38:00	2001-09-24 11:15:00	70	23	X2.6
21	2001-10-22 14:27:00	2001-10-22 15:08:00	2001-10-22 16:35:00	33	21	M6.7
22	2001-11-22 20:18:00	2001-11-22 20:36:00	2001-11-22 09:15:00	29	13	M9.9
23	2001-12-26 04:32:00	2001-12-26 05:40:00	2001-12-26 05:45:00	0	31	M7.1
24	2002-04-21 00:43:00	2002-04-21 01:51:00	2002-04-21 01:40:00	26	32	X1.5
25	2002-08-24 00:49:00	2002-08-24 01:12:00	2002-08-24 01:15:00	71	26	X3.1
26	2003-05-28 00:17:00	2003-05-28 00:27:00	2003-05-28 05:25:00	0	23	X3.6
27	2003-05-31 02:13:00	2003-05-31 02:24:00	2003-05-31 02:40:00	8	15	M9.3
28	2003-10-26 17:21:00	2003-10-26 18:19:00	2003-10-26 17:40:00	0	14	X1.2
29	2003-10-28 09:51:00	2003-10-28 11:10:00	2003-10-28 11:35:00	50	18	X17.2
30	2003-10-29 20:37:00	2003-10-29 20:49:00	2003-10-29 21:45:00	65	17	X10.0
31	2003-11-04 19:29:00	2003-11-04 19:50:00	2003-11-04 21:40:00	3	9	X28.0
32	2011-06-07 06:16:00	2011-06-07 06:41:00	2011-06-07 06:55:00	0	18	M2.5
33	2011-08-09 07:48:00	2011-08-09 08:05:00	2011-08-09 08:10:00	0	38	X6.9
34	2011-09-06 22:12:00	2011-09-06 22:20:00	2011-09-06 01:45:00	0	35	X2.1
35	2012-01-23 03:38:00	2012-01-23 03:59:00	2012-01-23 04:10:00	6	26	M8.7
36	2012-01-27 17:37:00	2012-01-27 18:37:00	2012-01-27 17:55:00	4	56	X1.7
37	2012-03-13 17:10:00	2012-03-13 17:12:00	2012-03-13 17:20:00	4	20	M7.9

Note. The tabular data modality includes the last three columns.

X-class flares indicating the highest energy release. Higher-intensity solar flares are more prone to producing high-energy-range SEP events. The last three columns of Table 2 describes the three tabular feature data used in this study. As shown in Figure 1, we collected tabular data from LMSAL, with SEP events sourced from the GSEP catalog and NSEP events from the HEK website.
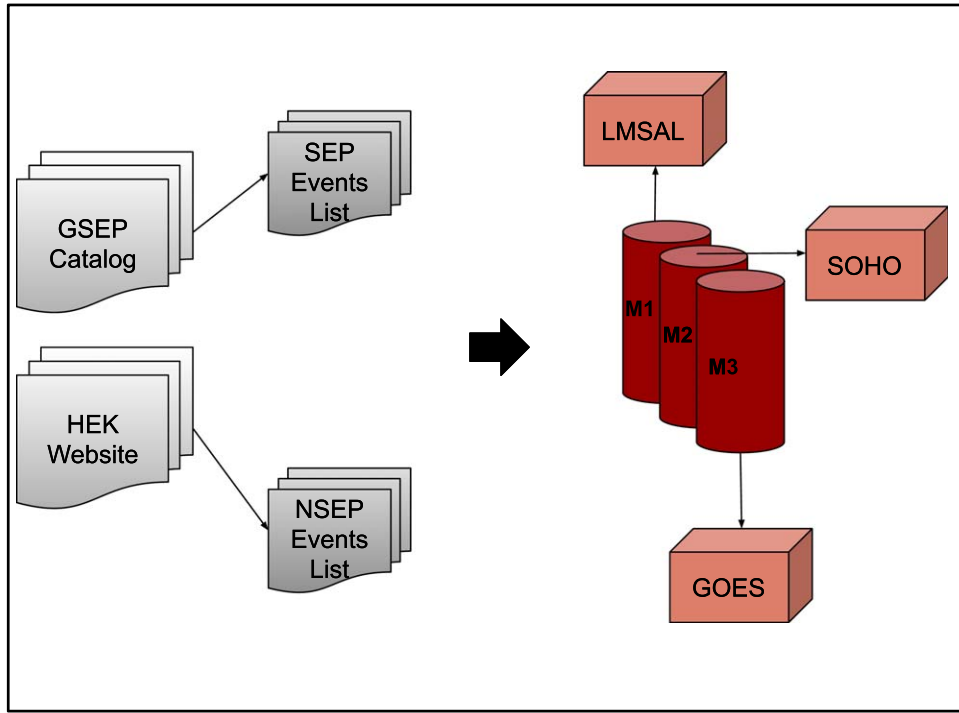
2.3. EUV Active Region Polygon Data

We employed EUV image data to capture spatial details of active regions. EUV images were retrieved from the Helioviewer platform, which provides access to data from the SOHO, specifically from the Extreme Ultraviolet Imaging Telescope (EIT) at the 304 Å wavelength. The data are publicly accessible.[3] We localized each parent AR of the solar flare events used in this study and segmented them using a minimum bounding rectangle (MBR). Figure 2 shows the full AR preprocessing pipeline. To generate the enclosed AR event

polygons, we began by cropping the high-resolution EIT images and resizing them to 200 by 200 pixels, followed by converting them to grayscale to emphasize intensity variations. We further refined this process by applying a widely used segmentation method that involves thresholding and contour detection (S. Suzuki & K. Abe 1985) to focus on the brightest regions within each cropped active region image. Specifically, we applied a binary threshold to highlight high-intensity areas, converting the image into a binary format where only the brightest pixels are retained. Following this, we used contour detection to identify the distinct boundaries of these bright regions, delineating the areas with significant activity. This segmentation isolates regions of interest based on intensity, allowing us to capture prominent structural features of active regions associated with solar events. A binary mask was then applied to highlight areas with significant X-ray intensity. This binary mask serves to filter out smaller, outlier segments, allowing only the largest AR segment to be retained for analysis. The identified edges were subsequently converted into polygonal shapes to capture the geometric structure of each active region accurately. Each processed image provides a

---

[3]   https://helioviewer.org/

Figure 1. Data collection pipeline. The figure shows the process for collecting SEP and non-SEP event data sets from the GSEP catalog, HEK website, and other sources. Three data modalities are included: M1 (tabular), M2 (EUV active region polygon), and M3 (GOES time-series proton flux).

(a) Original EUV image   (b) Edge detection   (c) Binary masking   (d) Polygon conversion
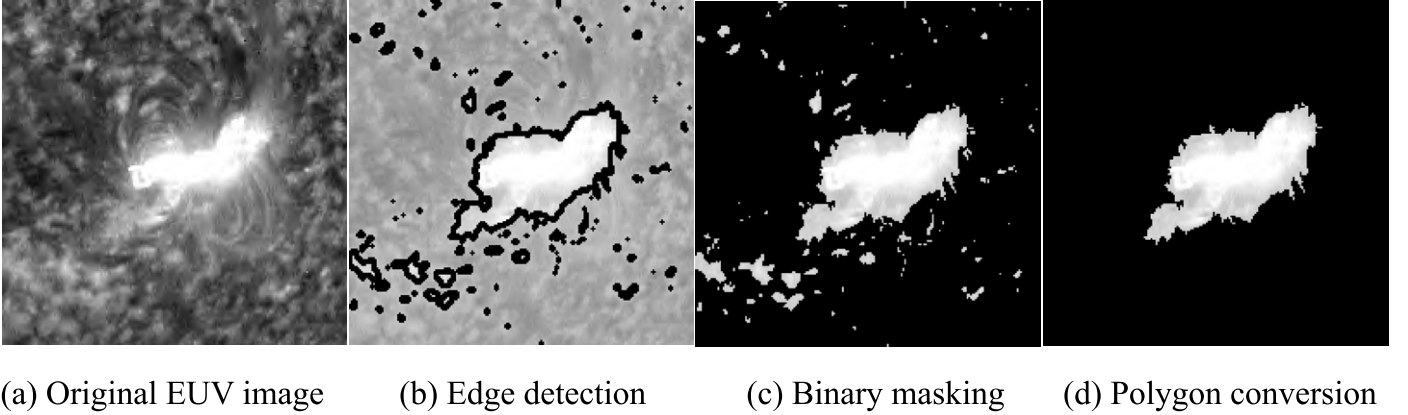
Figure 2. Preprocessing pipeline for EUV image data from the SOHO 304 Å wavelength.

snapshot taken just before the solar flare onset, encompassing data from both SEP and non-SEP events.

After extracting polygons, we employed three feature engineering techniques to enhance the predictive power of AR polygon data:

1. *Histogram of oriented gradients (HOG).* HOG captures the local gradient orientation, helping identify structural and intensity gradients within ARs. This method emphasizes edges and shapes, making it well suited for capturing spatial patterns related to SEP events (D. R. Dalal & B. Triggs 2005).

2. *Scale-invariant feature transform (SIFT).* SIFT detects key points and extracts features that are invariant to scale and rotation. It enhances the representation of AR regions by capturing distinct feature points, though it is less suited for stable images like solar images, where scale invariance is less critical (D. G. Lowe 2004).

3. *Attention mechanism (Attention).* The attention method focuses on specific image regions, emphasizing areas with high activity. By dynamically highlighting relevant regions, the model gains enhanced context, capturing nuanced intensity variations associated with solar activity (D. Bahdanau 2014; A. Vaswani 2017).

2.4. GOES Time-series Proton Flux Data

We retrieved the TS proton flux data from the GOES satellite through the NOAA National Centers for Environmental Information (NCEI).[4] We used data from the P6 channel, which measures proton flux in the energy range of 84–200 MeV. To capture variations in proton flux leading up

---

[4] https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/
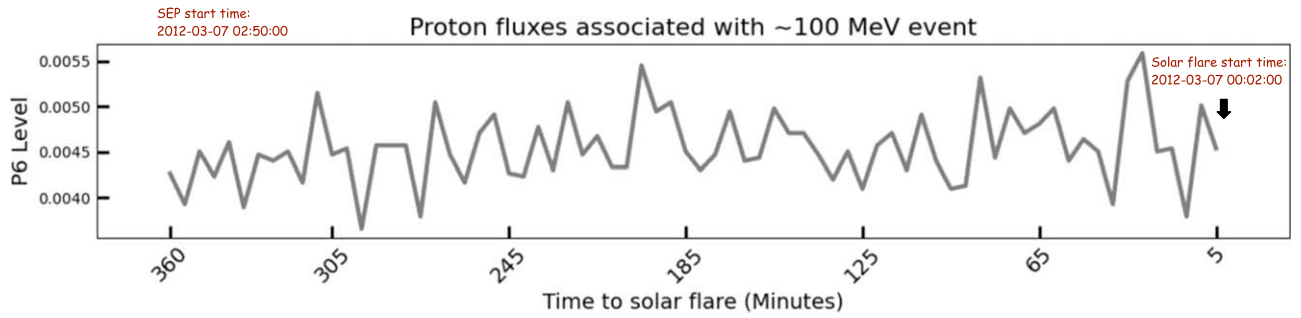
Figure 3. GOES time-series proton flux channel P6.

to SEP event onsets, we selected a 6 hr observation window prior to each parent SEP and Non-SEP events as an input data (P. Hosseinzadeh et al. 2024b). The P6 channel data provide valuable insights into the temporal dynamics of high-energy proton flux, allowing us to observe fluctuations that might be triggers to SEP events (P. Hosseinzadeh et al. 2024a). To improve comparability and focus on relevant flux changes, we applied Z-score normalization, as defined in Equation (1), which adjusts for background variations across the TS data:

$$z_i = \frac{x_i - \mu}{\sigma}, \tag{1}$$

where x_i is the raw proton flux value at time i, μ is the mean, and σ is the standard deviation of the 6 hr observation window. This process scales each data point by centering it around zero with a standard deviation of one. This approach highlights significant fluctuations by adjusting for background levels and enabling the detection of patterns that may signal SEP occurrences. Figure 3 illustrates a sample of P6 TS data, showing proton flux levels 6 hr before a solar flare starting at 2012 March 7 00:02:00, which subsequently leads to an SEP event at 2012 March 7 02:50:00.

2.5. t-SNE Visualization of Data Modalities

To illustrate the effectiveness of combining different data modalities, we applied t-distributed stochastic neighbor embedding (t-SNE) to visualize the separability of SEP and non-SEP events (L. Van der Maaten & G. Hinton 2008). Figure 4 demonstrates the clustering of SEP and non-SEP events using proton flux data alone (left) and the enhanced clustering achieved by integrating both EUV image data and proton flux data (right). The t-SNE plot shows that the combination of TS proton flux data and EUV features enhances the clustering of SEP events. The integration of different data sources allows the ML models to better differentiate between SEP and non-SEP events, providing a clearer separation in the high-dimensional feature space.

3. Methodology

In this section, we introduce the ML models that we used in this study to classify SEP and non-SEP events using different data modalities as input predictor(s). We will provide an overview of the models used, highlighting their general characteristics and their advantages in the SEP classification tasks. Figure 5 shows the architecture of all classifiers used in this work.

3.1. Random Forest

Random forest (RF) is an ensemble learning technique that trains multiple decision trees and combines their outputs to make a final prediction (L. Breiman 2001). This model operates by constructing each tree on a randomly selected subset of features and samples, which helps reduce overfitting by averaging out errors across the ensemble (T. K. Ho 1998). By aggregating results from multiple trees, a random forest lessens the influence of any one tree that may capture noise in the training data, thereby improving generalization (A. Liaw 2002). However, while this ensemble structure reduces overfitting compared to single decision trees, RF models can still overfit, in particular with highly complex or noisy data sets (T. Hastie 2009). The random forest model is particularly effective for high-dimensional data with complex interactions, as it can capture nonlinear relationships between features (D. R. Cutler et al. 2007).

3.2. Support Vector Machine

Support vector machine (SVM) is a widely used classifier, especially in high-dimensional spaces (C. Cortes & V. Vapnik 1995). The main principle of SVM is to find the optimal hyperplane that maximally separates the classes in the feature space. This hyperplane is determined by the support vectors, which are the data points closest to the decision boundary (C. J. Burges 1998). The use of kernel functions, such as the radial basis function, enables SVM to project the data into a higher-dimensional space where a linear separation can be achieved, even when the original data is not linearly separable (B. Schölkopf 2002). SVM is known for its strong theoretical foundation and its ability to perform well with a limited amount of data, making it a robust choice for classification tasks where clear boundaries exist between classes.

3.3. Logistic Regression

Logistic regression is a linear classifier that models the probability of an event occurring by fitting a logistic function to the input features (D. R. Cox 1958). In this study, logistic regression is applied only to tabular data. Despite its simplicity, logistic regression is a powerful baseline model because it provides a clear interpretation of how the features contribute to the likelihood of an event (D. W. Hosmer et al. 2013). The model assumes a linear relationship between the independent variables and the log-odds of the outcome, making it best suited for problems where the decision boundary is linear (J. M. Kernbach & V. E. Staartjes 2020). Although it may not capture more complex relationships between features, logistic regression is
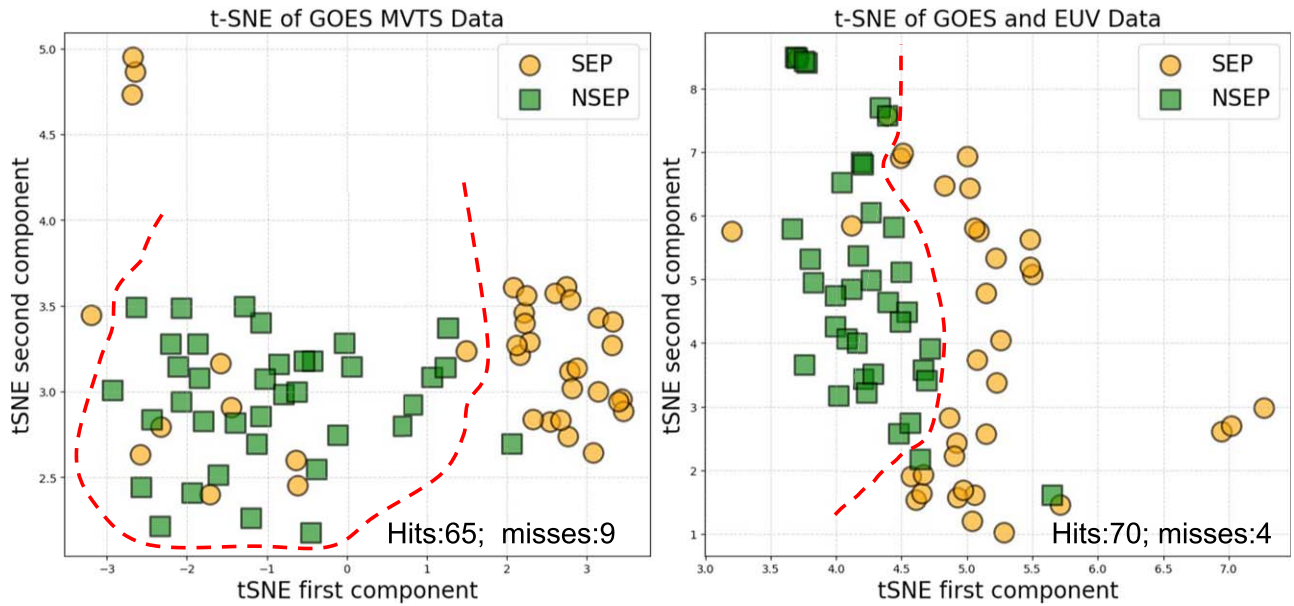
Figure 4. Two-dimensional t-SNE data projection using only GOES time-series data (left) and using image and GOES time-series data modalities.

computationally efficient and provides a solid baseline for comparison with more complex models.

3.4. XGBoost

XGBoost is a highly efficient gradient-boosting algorithm that constructs decision trees in a sequential manner (T. Chen & C. Guestrin 2016). Each new tree corrects the errors made by the previous trees, which allows the model to improve its accuracy progressively (J. H. Friedman 2001). XGBoost is particularly well suited for large data sets and high-dimensional data, as it includes various optimizations that speed up the training process. Although XGBoost can be prone to over-fitting, especially in complex or noisy data, it includes regularization techniques that help mitigate this risk and enhance its generalization to unseen data. This model is known for its superior performance in ML competitions and is widely recognized for its ability to handle both numeric and categorical features effectively.

3.5. CatBoost

CatBoost is another gradient-boosting algorithm that specializes in handling categorical features (A. V. Dorogush et al. 2018). One of the key strengths of CatBoost is its ability to process categorical data directly, without the need for extensive preprocessing such as one-hot encoding. This allows the model to preserve the natural relationships within categorical features, reducing the risk of overfitting. Unlike XGBoost, CatBoost uses ordered boosting, which minimizes target leakage by training with past observations only, thereby reducing the risk of overfitting while preserving natural relationships within categorical features (L. Prokhorenkova et al. 2018).

3.6. Multilayer Perceptron

The multilayer perceptron (MLP) is a type of feed-forward neural network that consists of an input layer, one or more hidden layers, and an output layer (I. Goodfellow 2016). Each layer is composed of neurons that apply a nonlinear activation
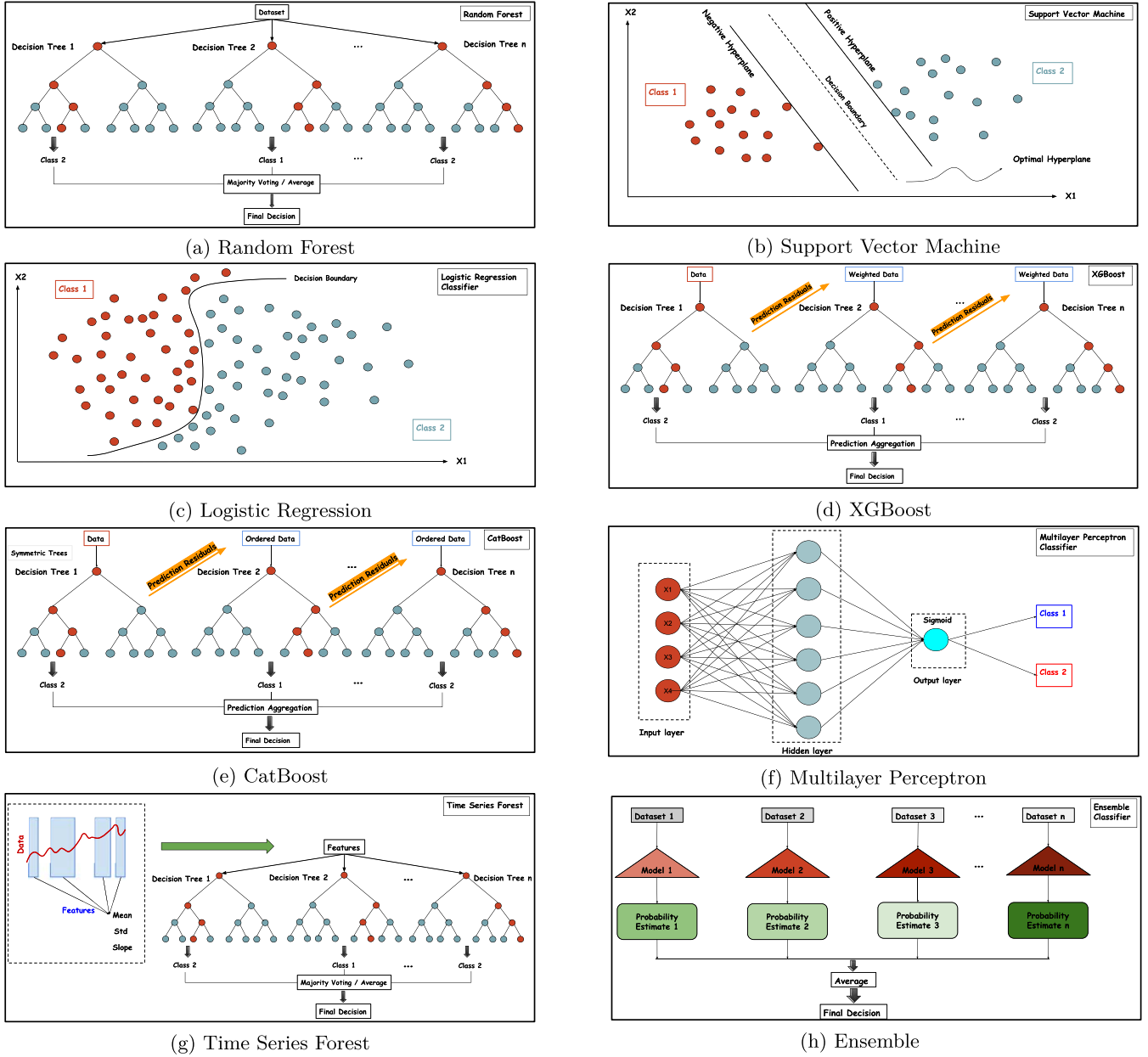
function to the inputs from the previous layer, allowing the network to learn complex, nonlinear relationships in the data (C. C. Aggarwal 2018). An MLP is a mathematical function mapping a set of input values to output values (I. Goodfellow 2016). MLPs are particularly well suited for tasks involving high-dimensional and nonlinear data, as they can capture intricate relationships within the input features. By stacking multiple hidden layers, an MLP can model intricate patterns in the data, making it a powerful tool for classification tasks that require deep representations (J. Schmidhuber 2015).

3.7. Time-series Forest

Time-series forest (TSF) is an ensemble learning method designed specifically for TS classification (H. Deng et al. 2013). TSF constructs decision trees by selecting random intervals from the TS and calculating summary statistics such as mean, variance, and slope over these intervals. These summary statistics are then used as features to classify the TS (J. Lines et al. 2016). TSF is particularly effective for TS data because it captures both local and global patterns, allowing it to identify temporal trends and variations that may be indicative of SEP events (A. Bagnall et al. 2017). The random selection of intervals helps the model avoid overfitting while maintaining good predictive performance.

3.8. Ensemble Learning

In this study, we employ several ensemble learning strategies by combining models trained on different combinations of data modalities, including TS, image-based, and tabular data. Each ensemble model integrates the predictions of individual classifiers trained on specific modalities, and we average the predicted probabilities (probability estimates) to generate the final classification decision (J. Kittler et al. 1998). We explore different ensemble configurations depending on the data modalities combined. For example, one ensemble model combines TS and image-based data, another combines TS and tabular data, and a third ensemble integrates all three modalities. By averaging the predicted probabilities from the classifiers

(a) Random Forest

(b) Support Vector Machine

(c) Logistic Regression

(d) XGBoost

(e) CatBoost

(f) Multilayer Perceptron

(g) Time Series Forest

(h) Ensemble

Figure 5. Architecture of the classifiers used in this study: (a) random forest, (b) support vector machine, (c) logistic regression, (d) XGBoost, (e) CatBoost, (f) multilayer perceptron, (g) time-series forest, and (h) ensemble.

trained on each modality, each ensemble model leverages the strengths of its respective data sources to enhance predictive accuracy and robustness. This approach ensures that our ensemble models provide more comprehensive and reliable predictions compared to individual models, as each combination of data modalities contributes unique insights, ultimately improving performance in the classification of SEP events.

In this work, we used four ensemble models to evaluate the multimodality effects of the SEP prediction:

1. *TS-HOG* combines TS data (using TSF) and active region polygon data (using RF with HOG features).
2. *TS-TB* integrates TS data (using TSF) and tabular data (using SVM).
3. *HOG-TB* combines polygon data (using RF with HOG features) and tabular data (using SVM).

4. *TS-HOG-TB* incorporates all three modalities (TS, polygon, and tabular data, using the best methods for each modality).

3.9. Evaluation Metrics

In this section, we outline the metrics used to evaluate the effectiveness of the SEP classification models in our study. To measure classification performance, we used two primary metrics: accuracy and F1-score. Accuracy represents the overall proportion of correct predictions made by the model. To better understand the F1-score, we first define precision and recall. Precision is the ratio of correctly predicted positive cases to the total predicted positive cases. This metric indicates how many of the events identified by the model as positive are actually
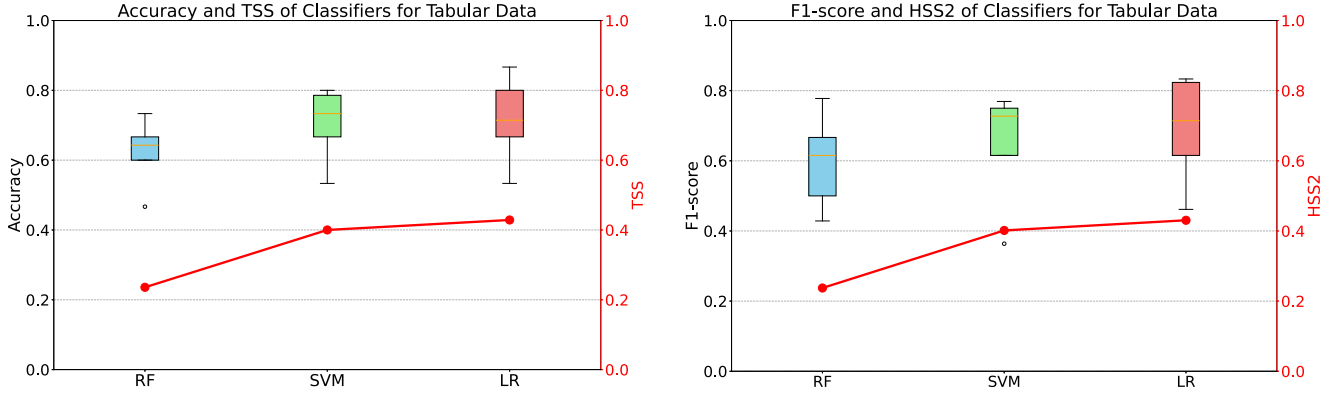
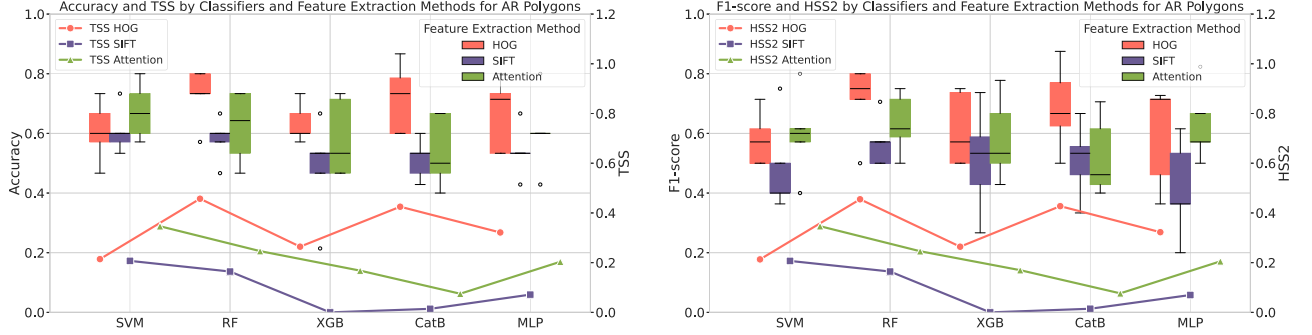**Figure 6.** Performance of classifiers for tabular data.

Figure 7. Performance of classifiers and feature extraction methods for EUV active region polygons.

correct, minimizing false positives. Recall, on the other hand, is the ratio of correctly predicted positive cases to the total actual positive cases. This metric measures the model's ability to capture true positive cases, minimizing false negatives. The F1-score, combines precision and recall into a single metric, offering a balanced view of the model's ability to identify positive cases while controlling for both false positives and false negatives. In addition to conventional ML metrics, we also employ the true skill statistic (TSS) and the updated Heidke skill score (HSS2), which are commonly used in solar event prediction tasks and are less sensitive to class balanced and imbalanced problems.

4. Experimental Results

In this section, we present the results of the experiments performed on the three different data modalities: tabular data, AR polygon data (extracted from EUV image data), and TS proton flux data. Each modality was used to train different ML models, and the best-performing methods for each modality were identified and later used in ensemble models. We used fivefold cross validation to assess all the model performances. This approach divides the data set into five subsets/folds. For each iteration, one fold is reserved as the validation set while the remaining four folds are used for training.

Figure 6 shows the classification results for the tabular data, which includes features such as sunspot counts, active region counts, and solar flare classifications. Among the models tested, the support vector machine model achieved the highest accuracy and F1-score. The strong performance of SVM can be attributed to its ability to create an optimal decision boundary in high-dimensional spaces, which is crucial for distinguishing between SEP and non-SEP events based on tabular features. RF

achieved the lowest performance levels, likely due to its complex architecture compared to the relatively low dimensionality of the tabular features. Logistic regression, while computationally efficient and interpretable, provided lower average accuracy, due to its assumption of linear relationships between features and outcomes, limiting its capacity to capture the complex dynamics in the data. Although logistic regression shows slightly better TSS and HSS2 scores, it has higher variances in accuracy and F1-score, indicating lower model robustness. Figure 7 introduces the performance of the five models that we trained on the engineered data from the EUV images. The figure shows that the HOG feature extraction method achieves the best classification results compared to other feature engineering methods (e.g., SIFT) across most of the models. HOG's ability to capture local gradient orientations enhances its capacity to represent critical edge and shape information, which is essential for identifying spatial patterns in ARs associated with SEP events. This method proves particularly effective in combination with RF, which stands as the highest-performing classifier among all tested combinations. The success of HOG-RF highlights how HOG's focus on intensity gradients aligns well with the structural characteristics of active regions, providing a more accurate basis for distinguishing SEP-related features.

In contrast, the SIFT method achieved the lowest performance, due to the characteristics of the generated features that are invariant to scale and rotation. The stability of solar images may diminish the effectiveness of SIFT's invariance properties, making it less well suited for this task. Similarly, attention-based methods, while capable of dynamically focusing on important regions of the input data, are not able to capture the prominent patterns useful for prediction. While attention methods coupled with certain classifiers (e.g., MLP and
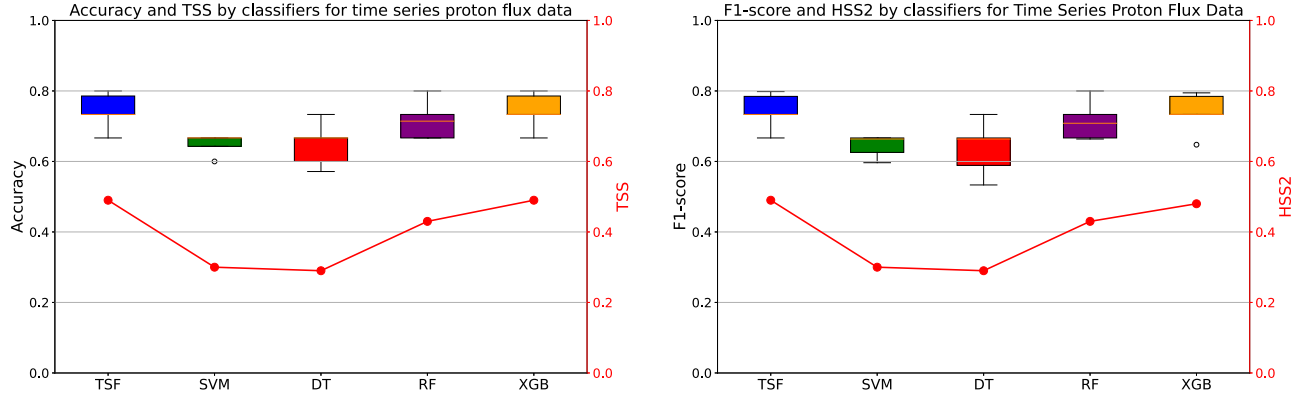
Figure 8. ML models' performance when trained on time-series proton flux data from the GOES P6 channel.
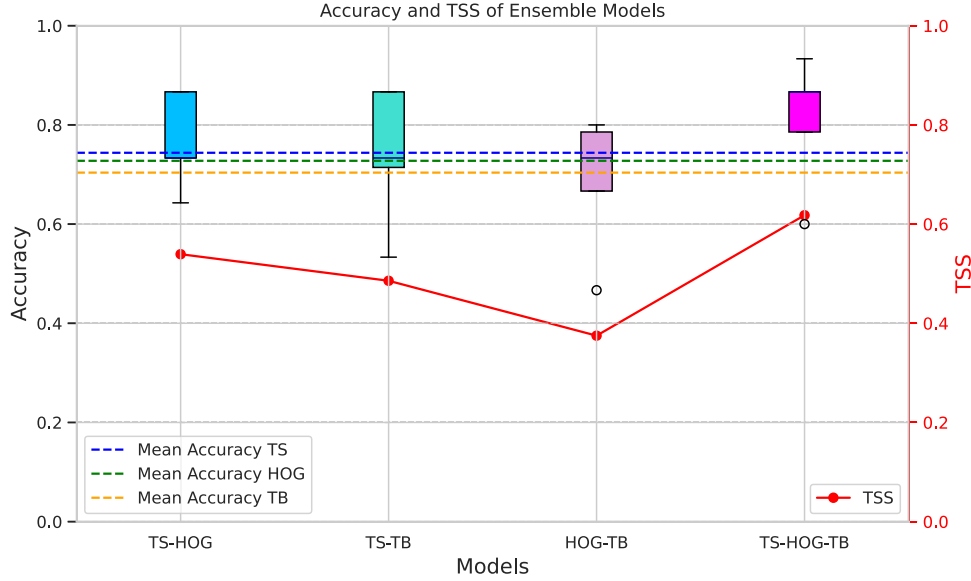
Figure 9. Plots showing the accuracy and TSS of ensemble models. Dashed lines represent mean accuracies of unimodal models: TSF, RF, and SVM.

SVM) demonstrate competitive accuracy, they lack the precision of HOG-RF, which is particularly adept at capturing spatial variations essential for accurate SEP prediction.

Figure 8 shows the performance levels of five ML models when trained on TS proton flux data from the GOES P6 channel using the accuracy, F1-score, TSS, and HSS2 metrics. Among the tested models, TSF outperformed the others by effectively leveraging both short-term fluctuations and long-term trends in the TS data. By extracting summary statistics, such as mean and variance, from random intervals of the TS, TSF effectively models both global and local patterns. The ability of TSF to handle different temporal scales is particularly well suited for the task of SEP event classification. While RF and XGBoost also performed well, they did not achieve the highest accuracy, due to their reliance on fixed tree structures, which limits their ability to capture the intricate, multiscale temporal patterns necessary for distinguishing SEP events.

Finally, Figure 9 presents the results of the ensemble models that combine the best-performing methods from each modality. The results show that TS-HOG-TB, which integrates all three modalities, provided the highest accuracy and TSS score. This is expected, as combining TS, spatial, and tabular data offers the most comprehensive representation of the conditions leading to SEP events. By leveraging both temporal patterns

in proton flux data and spatial information from active region polygons, the TS-HOG-TB model has a comprehensive view of solar condition, enabling it to better distinguish between SEP and non-SEP events. TS-HOG, which combines TS and EUV polygon data, also achieved strong results, reinforcing the importance of integrating temporal and spatial features for SEP prediction. TS-TB, which integrates TS and tabular data, performed slightly lower than the previous two ensembles, suggesting that spatial features play a critical role in improving classification accuracy. HOG-TB, which combines polygon and tabular data, achieved the lowest accuracy among the ensembles, likely due to the absence of temporal information, which is crucial for capturing dynamic changes in proton flux leading up to SEP events.

The results highlight the importance of multimodal data integration in SEP prediction tasks. The ensemble models that took all the data modalities as input consistently outperformed single-modality models. This shows that each modality provides unique and complementary insights into solar activity. By combining the strengths of each modality, the ensemble models offer a more robust and accurate prediction of SEP events.

To show the performance of the best-performing ensemble model (TS-HOG-TB), we evaluate both accuracy and TSS
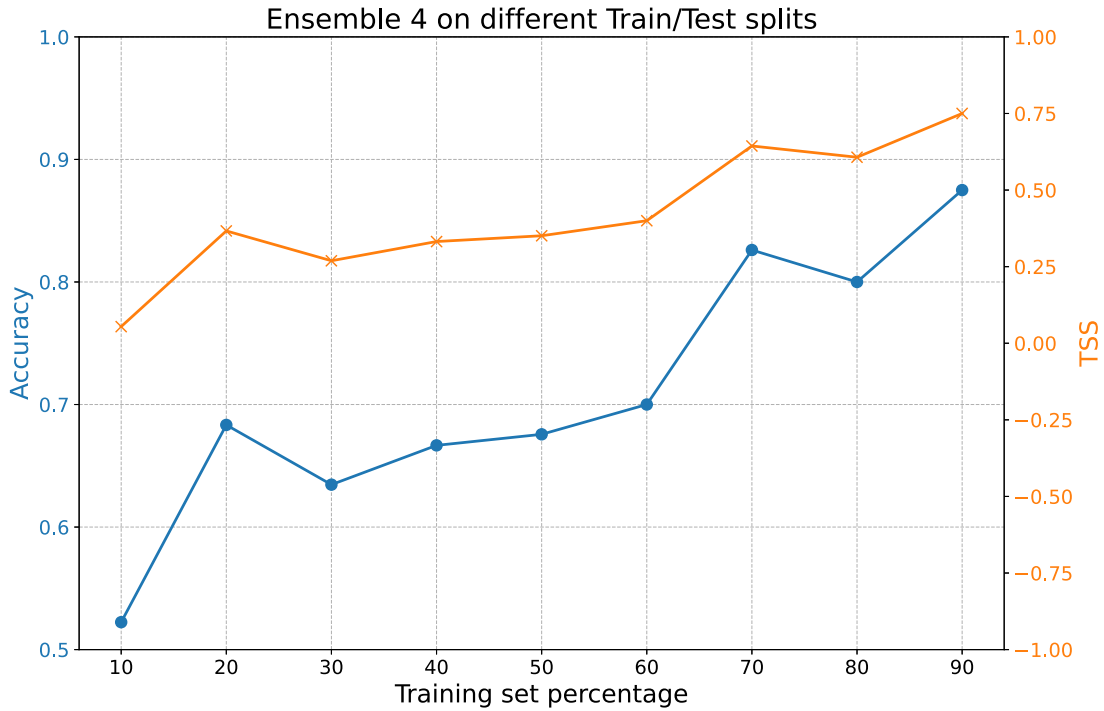
Figure 10. Performance of TS-HOG-TB (in terms of accuracy and TSS) on the train/test partitions.

across different train/test splits. Figure 10 shows the results of the effect of training data volumes of the testing accuracy of TS-HOG-TB. The plot shows that TS-HOG-TB maintains a high level of accuracy even as the training set size decreases, demonstrating its robustness in predicting SEP events. This highlights the ensemble model's ability to generalize well with a fixed train/test split, while also preserving high accuracy and TSS scores. Additionally, The plot suggests that increasing the amount of training data leads to marginal gains in both accuracy and TSS, indicating that TS-HOG-TB can still benefit from larger training sets. However, the model achieves near-optimal performance with approximately 60% of the data, suggesting a good balance between training data requirements and predictive performance. These results confirm the importance of integrating multimodal data in SEP event prediction and demonstrate the potential of TS-HOG-TB for operational use in space-weather forecasting, using both cross validation and train/test split evaluations effectively.

The last experiment we performed consisted of a sensitivity analysis to evaluate the robustness of our best-performing ensemble model (TS-HOG-TB) under noisy test conditions. We used Gaussian noise with a mean of 0 and a standard deviation of 0.5, which was separately introduced to the unimodal GOES TS-based model, the unimodal EUV polygon-based model, and directly to TS-HOG-TB. Additionally, we compare these results to the baseline performance of TS-HOG-TB without added noise, allowing for a clear assessment of the impact of noise on model accuracy. Figure 11 shows the results of the analysis that indicated that TS-HOG-TB, even under the stress of significant noise, maintains a higher level of accuracy than the individual models when they face similar noisy conditions. This not only shows the ensemble's ability to dampen the negative effects of noise through its integrated multimodal strategy but also highlights its superior capability in utilizing diverse sources of information to maintain prediction accuracy. Moreover, despite the presence of noise,

TS-HOG-TB's outputs are markedly more accurate than those of the TS-based and EUV polygon-based models under similar conditions. We also note that, while all models experience some accuracy decline due to noise, TS-HOG-TB model is least affected, suggesting that its structure effectively mitigates the impact of data perturbations. The ensemble has a high ability to aggregate and harmonize insights from its constituent models, enabling TS-HOG-TB to outperform any single model in isolation.

The ability of TS-HOG-TB to maintain a high level of performance, despite substantial noise, underscores the effectiveness of using advanced ensemble methods in predictive modeling. By integrating information from multiple sources, TS-HOG-TB not only improves its overall accuracy but also enhances its stability and reliability, making it an exemplary model for real-world applications where data quality is variable and often unpredictable.

5. Discussion

In this section, we analyze the model performance in predicting SEP events, examining solar flare class distribution, solar cycle patterns, and the effect of TS length on prediction accuracy.

Figure 12 illustrates the flare class distribution, offering insights into the occurrence of solar flares across different classes: C, M, and X. In this study, we focus on ∼100 MeV SEP events that may influence the representation of flare classes in this plot, highlighting instances associated with higher-intensity flares. While the plot suggests a relatively higher representation of X-class flares among the observed 100 MeV SEP events, this selection bias could potentially amplify the visibility of these events. In reality, C-class flares are more frequently observed in solar activity. However, Figure 12 only shows that SEPs are more correlated with strong flares. We coupled this information with solar cycle data to
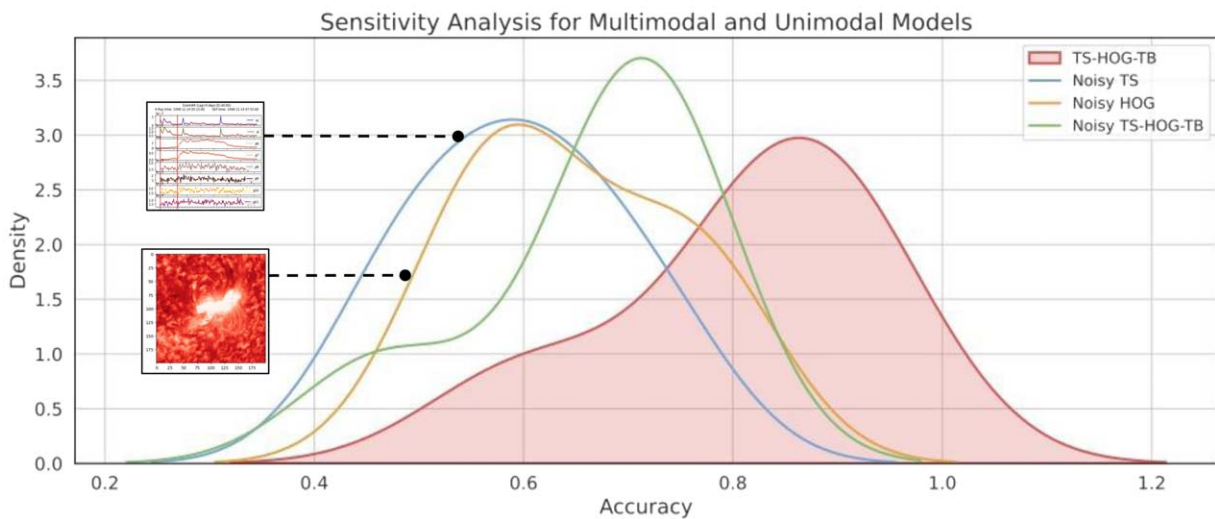
Figure 11. Sensitivity analysis showing the impact of noise on the performance of TS-HOG-TB and individual models. Note that the curves represent Gaussian kernel density estimates of accuracy values, which are smoothed representations of the data. While the kernel density estimate curves may extend slightly beyond the range of [0, 1] due to the smoothing process, all observed accuracy values remain within the valid range of [0, 1].
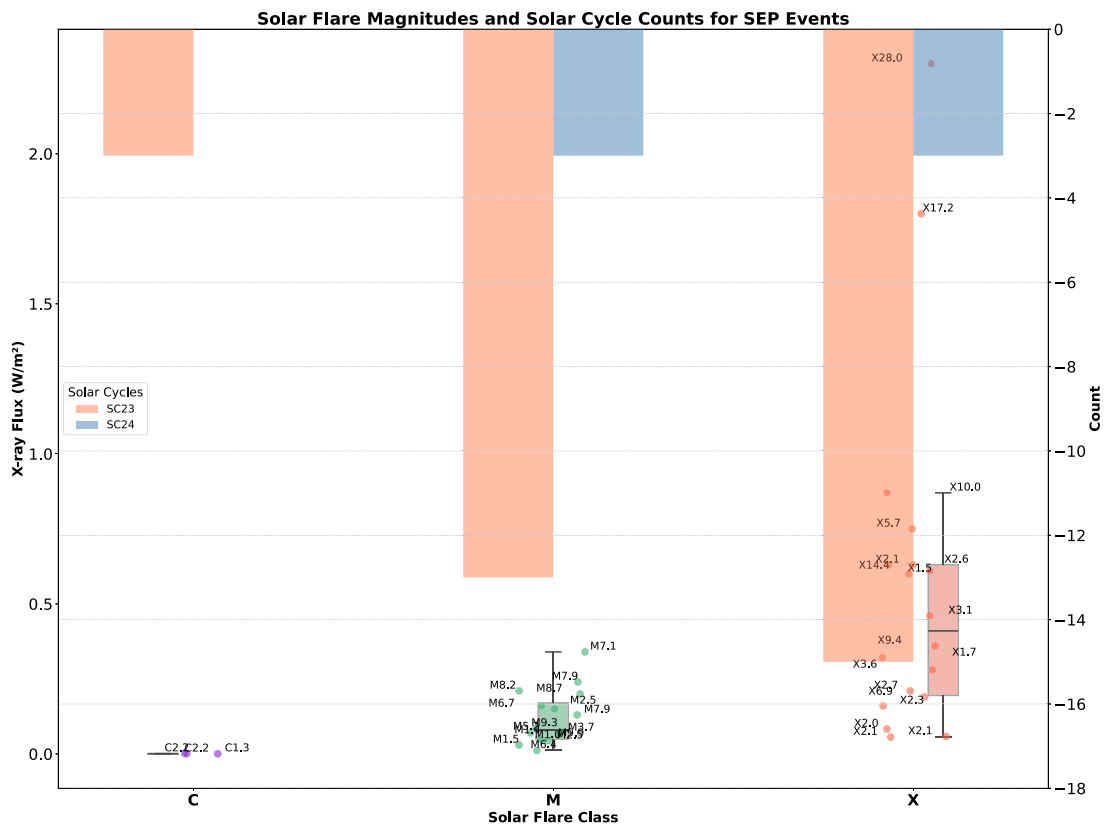

Figure 12. SEP X-ray flux and flare class distribution, and their corresponding solar cycle counts.

gain a deeper understanding of the relationship between solar activity and SEP events. The bar plots show the number of SEP events occurring within each solar cycle for each solar flare class. This analysis provides evidence of the cyclical nature of solar activity and its direct impact on the SEP prediction problem. The figure shows that, among SEP events, higher-intensity flares (M- and X-class) with elevated X-ray flux levels are more prominent, while lower-intensity C-class flares are less common. This distribution indicates that SEP events are

more frequently associated with intense flares. Although SEP events seem to cluster around the peaks of the solar cycles, the limited data set prevents us from definitively concluding that a strong link exists between SEP occurrence and solar cycle phases. Nonetheless, the observed trends suggest that periods of high solar activity and elevated X-ray flux may increase the likelihood of SEP events.

Furthermore, we analyzed the performance of the TS-HOG-TB model across four distinct EUV channels: 171, 195, 284,
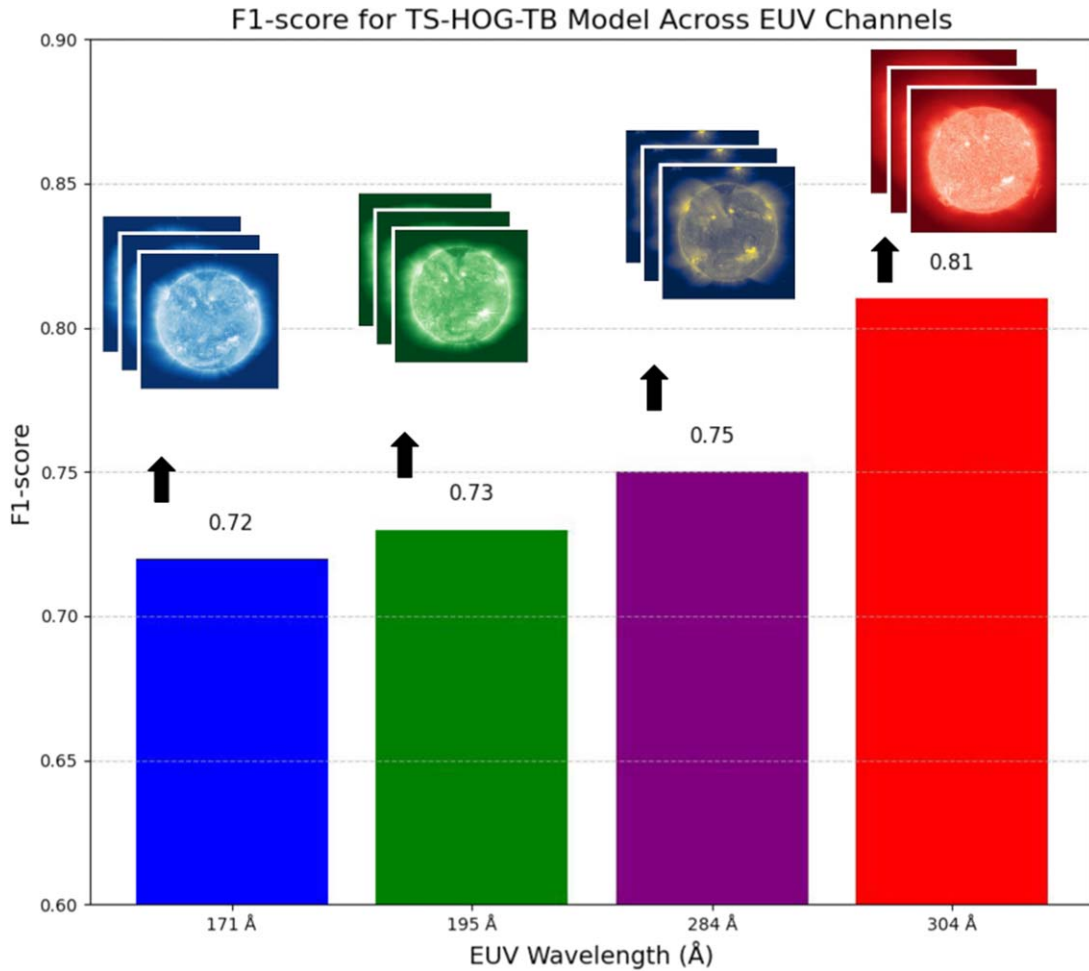
Figure 13. F1-scores for the TS-HOG-TB model across four EUV channels. The F1-scores are displayed for the 171, 195, 284, and 304 Å channels, corresponding to distinct layers and temperature regimes in the solar atmosphere.

and 304 Å, each capturing unique solar features linked to specific temperature regimes. As shown in Figure 13, the 304 Å channel from SOHO's EIT achieves the highest F1-score (0.81), underscoring its strong predictive capability for SEP events. Additionally, the 284 Å (F1-score: 0.75) and 195 Å (F1-score: 0.73) channels exhibit competitive performance, reflecting their ability to capture features from the hotter regions of the solar corona. In contrast, the 171 Å channel shows the lowest F1-score (0.72), which may be attributed to its emphasis on cooler coronal structures that offer less predictive information for SEP events. These results highlight the predictive superiority of the 304 Å channel, which was primarily utilized in our study due to its ability to provide the most informative features for the ensemble TS-HOG-TB model.

We investigated the influence of varying TS data lengths on the performance of TS-HOG-TB, which has been identified as our best-performing model. As shown in Figure 14, the sequence analysis provides an evaluation of the model across different observational window lengths for GOES proton flux, ranging from 1 to 9 hr. The line plots in Figure 14 show both the average TSS and the F1-score, indicating a general trend of increased predictive performance with longer observation windows. This trend is especially noticeable after the 5 hr mark, where both the TSS and F1-score see significant

improvements. This suggests that the model benefits from a more extended temporal context, likely due to the richer information content that allows for more accurate detection and classification of SEP events. However, the performance gain starts to plateau around the 7 hr mark. This plateau indicates a diminishing return on prediction improvement with longer TS data, suggesting an optimal balance between data quantity and model performance within the 6–7 hr range.

We have only analyzed the balanced setting in the article so far. Here, we extend our analysis to include an imbalanced setting, which reflects a more realistic scenario for SEP prediction. The imbalanced data set comprises 141 total samples, with 104 non-SEP samples and 37 SEP samples. Figure 15 provides a comparative evaluation of the TS-HOG-TB model's performance under both balanced and imbalanced conditions. The left subfigure displays the accuracy box plot for each setting, along with the mean precision plotted as a line beneath. The right subfigure shows the F1-score box plot for each setting, with the mean recall displayed as a line plot below. This dual-visualization strategy enables a clear comparison of central tendency and variability in the model's performance for both data set configurations. The balanced data set exhibits higher median accuracy and F1-score, along with more compact interquartile ranges, reflecting consistent model behavior. Conversely, in the imbalanced setting, there is greater dispersion in the accuracy and F1-score distributions,
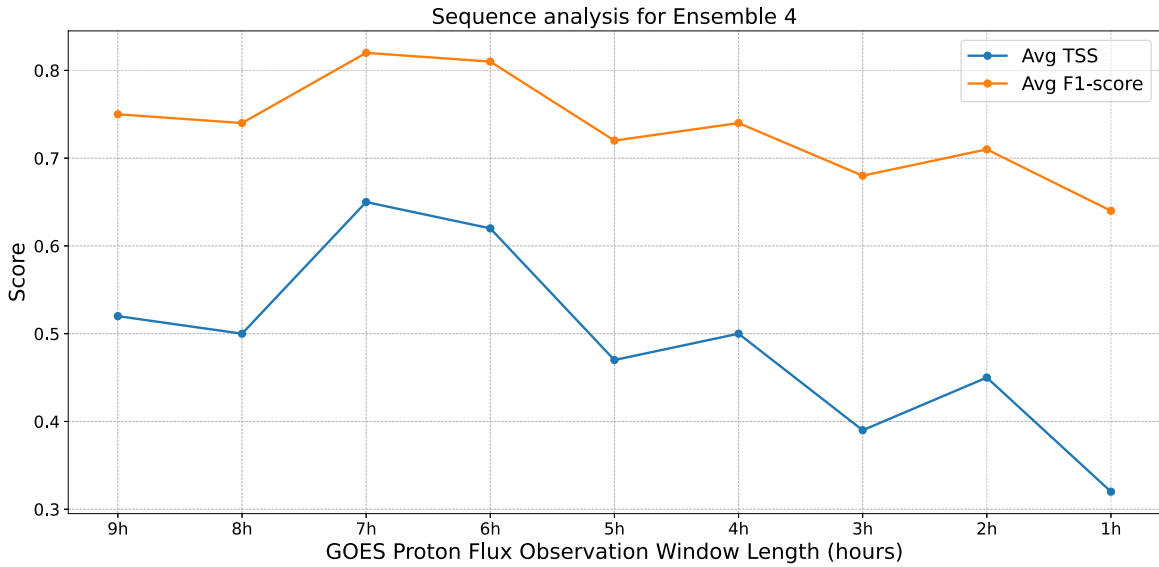
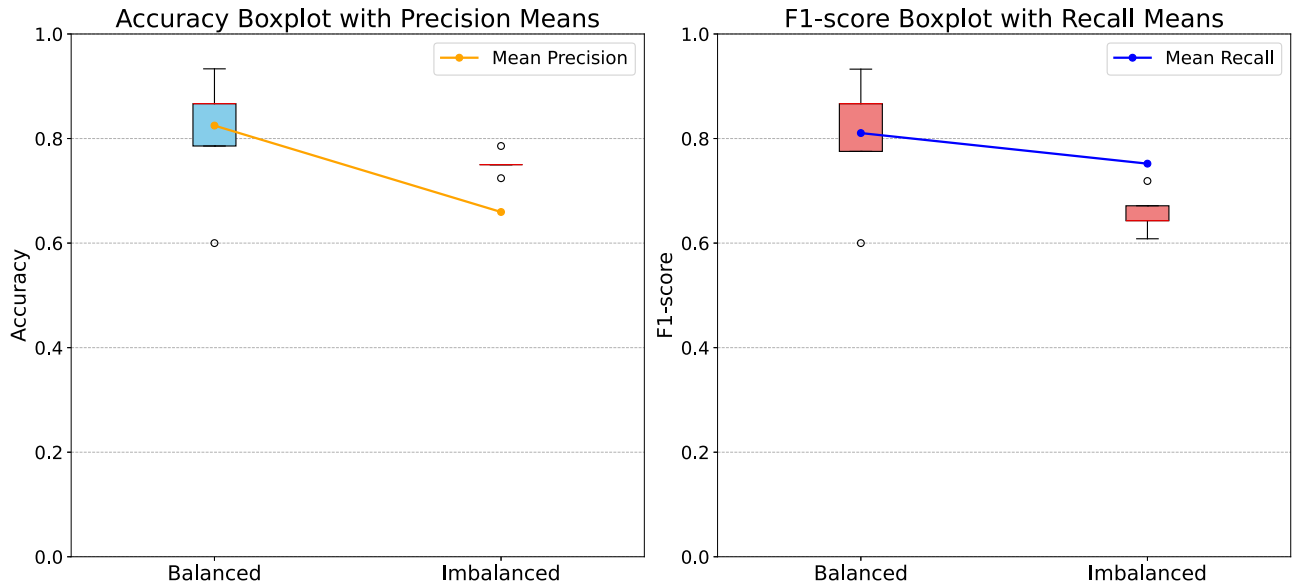**Figure 14.** Sequence analysis for TS-HOG-TB model.

Figure 15. Performance results of TS-HOG-TB for balanced and imbalanced settings.

highlighting the model's difficulty in handling the uneven class distribution. Precision drops significantly in the imbalanced data set, while recall remains comparatively stable, indicating that the model becomes more conservative in declaring SEP events, to reduce false positives.

Given the imbalanced nature of the SEP classification problem, recall emerges as the most critical metric for evaluating the model's performance. The objective is to correctly identify as many SEP events as possible, since missed SEP events (false negatives) can have severe operational and safety consequences for space missions, satellite integrity, and astronaut health. While precision captures how often predicted SEP events are correct, a slight increase in false positives (predicting SEP when it is actually non-SEP) is far less consequential than missing a true SEP. The stability of recall across both the balanced and imbalanced settings, as shown in Figure 15, reinforces its significance as the primary evaluation metric. By maintaining a higher and more consistent recall, the model shows its capacity to detect SEP events

reliably, even when faced with an imbalanced data set. However, we acknowledge that, while thousands of non-SEP samples exist, we only analyzed 104 of them for the imbalanced setting in this study. Collecting more non-SEP samples is a time-consuming process, as it requires manual collection, trimming, and processing of the data. Expanding the analysis to incorporate a larger pool of non-SEP samples remains an important direction for future work.

Finally, although direct comparison with other works in the SEP prediction space is challenging due to differences in data sets and experimental setups, we have compared our best model with several recent studies. For instance, the MEM-PSEP-I model (S. Chatterjee et al. 2024) reported an accuracy of 0.80, which is comparable to ours, though their TSS score is higher. Similarly, Lavasa et al. (E. Lavasa et al. 2021) achieved a recall of 0.76 using their best model, a random forest, whereas our ensemble TS-HOG-TB model achieved a recall of 0.81. However, their TSS score of 0.75 was higher than ours. Additionally, A. Ali et al. (2024), who focused on predicting

the occurrence of SPEs using ML, reported a TSS of 0.68 using XGBoost, along with a recall of 0.90. These comparisons highlight the competitiveness of our approach while also emphasizing the variability in metrics due to differences in data, model choices, and experimental design.

6. Conclusion

In this study, we developed an ensemble ML model that integrates multiple data modalities (solar EUV image, GOES TS proton flux, and tabular data) to predict high-impact SEP events. The ensemble model demonstrates improved prediction accuracy for high-energy SEP events and exhibits robustness against noise typically present in observational data. A key aspect of our approach is the extraction of active region polygons from solar EUV images, which captures spatial information on solar activity regions and contributes significantly to model accuracy by highlighting regions of heightened solar activity. Our analyses indicate that ensemble models outperform those trained on single data modalities, in particular in environments that mimic operational conditions, emphasizing the strength of multimodal data integration. This approach allows for a more comprehensive representation of solar conditions, enhancing the model's resilience and predictive stability across different phases of solar activity. Additionally, our investigation into the impact of TS length reveals that longer observation windows lead to improved performance up to a certain threshold, after which the predictive gains plateau, suggesting an optimal balance between data quantity and model efficiency in the 6–7 hr range. These findings underscore the value of integrating temporal, spatial, and tabular data for robust SEP prediction, and they highlight the potential of our approach to support early-warning systems with greater accuracy and reliability. Future work will focus on real-time data integration and new ML techniques to further enhance the predictive capabilities and operational applicability of SEP forecasting models.

ORCID iDs

Pouya Hosseinzadeh ⬤ https://orcid.org/0000-0001-8045-2709
Soukaina Filali Boubrahimi ⬤ https://orcid.org/0000-0001-5693-6383
Shah Muhammad Hamdi ⬤ https://orcid.org/0000-0002-9303-7835

References

Aggarwal, C. C. 2018, Neural Networks and Deep Learning (Cham: Springer)
Ali, A., Sadykov, V., Kosovichev, A., et al. 2024, ApJS, 270, 15
Aminalragia-Giamini, S., Raptis, S., Anastasiadis, A., et al. 2021, JSWSC, 11, 59
Aran, A., Sanahuja, B., & Lario, D. 2006, AdSpR, 37, 1240
Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. 2017, Data Min. Knowl. Discov., 31, 606
Bahdanau, D. 2014, arXiv:1409.0473
Boubrahimi, S. F., Aydin, B., Martens, P., & Angryk, R. 2017, in 2017 IEEE Int. Conf. on Big Data (Big Data) (Piscataway, NJ: IEEE), 2533
Breiman, L. 2001, Mach. Learn., 45, 5
Burges, C. J. 1998, Data Min. Knowl. Discov., 2, 121
Camporeale, E. 2019, SpWea, 17, 1166
Carrington, R. C. 1859, MNRAS, 20, 13
Chatterjee, S., Dayeh, M. A., Muñoz-Jaramillo, A., et al. 2024, SpWea, 22, e2023SW003568
Chen, T., & Guestrin, C. 2016, in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, ed. B. Krishnapuram (New York: Association for Computing Machinery), 785
Cliver, E. W., & Svalgaard, L. 2004, SoPh, 224, 407
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Cox, D. R. 1958, J. R. Stat. Soc. B, 20, 215
Cutler, D. R., Edwards, T. C., Beard, K. H., et al. 2007, Ecology, 88, 2783
Dalal, D. R., & Triggs, B. 2005, in 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05) (Piscataway, NJ: IEEE), 886
Davies, E. E., Forsyth, R. J., Winslow, R. M., Möstl, C., & Lugaz, N. 2021, ApJ, 923, 136
Deng, H., Runger, G., Tuv, E., & Vladimir, M. 2013, Inform. Sci., 239, 142
Desai, M., Mason, G., Dayeh, M., et al. 2016, ApJ, 828, 106
Dorogush, A. V., Ershov, V., & Gulin, A. 2018, arXiv:1810.11363
Filali Boubrahimi, S., Neema, A., Nassar, A., Hosseinzadeh, P., & Hamdi, S. M. 2024, WRR, 60, e2023WR036342
Forbush, S. E. 1946, PhRv, 70, 771
Friedman, J. H. 2001, AnSta, 29, 1189
Furukawa, S., Nagamatsu, A., Nenoi, M., et al. 2020, BioMed Res. Int., 2020, 4703286
Goodfellow, I. 2016, Deep Learning (Cambridge, MA: MIT Press)
Gopalswamy, N., Xie, H., Akiyama, S., Mäkelä, P. A., & Yashiro, S. 2014, EP&S, 66, 104
Gopalswamy, N., Yashiro, S., Michałek, G., et al. 2002, ApJL, 572, L103
Hastie, T. 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Berlin: Springer)
Hathaway, D. H. 2015, LRSP, 12, 4
Ho, T. K. 1998, ITPAM, 20, 832
Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. 2013, Applied Logistic Regression (New York: Wiley),
Hosseinzadeh, P., Bahri, O., Li, P., Boubrahimi, S. F., & Hamdi, S. M. 2023, in 2023 Int. Conf. on Machine Learning and Applications (ICMLA) (Piscataway, NJ: IEEE), 1248
Hosseinzadeh, P., Boubrahimi, S. F., & Hamdi, S. M. 2024a, ApJS, 270, 31
Hosseinzadeh, P., Boubrahimi, S. F, & Hamdi, S. M. 2024b, SpWea, 22, e2024SW003982
Hosseinzadeh, P., Li, P., Bahri, O., Boubrahimi, S. F., & Hamdi, S. M. 2024c, in 2024 IEEE Int. Conf. on Big Data (BigData) (Piscataway, NJ: IEEE), 1327
Jackson, I., & Martens, P. 2024, ApJS, 272, 37
Kahler, S. 1982, JGR, 87, 3439
Kasapis, S., Kitiashvili, I. N., Kosovich, P., et al. 2024, ApJ, 974, 131
Kasapis, S., Zhao, L., Chen, Y., et al. 2022, SpWea, 20, e2021SW002842
Kernbach, J. M., & Staartjes, V. E. 2020, arXiv:2006.15069
Kittler, J., Hatef, M., Duin, R. P., & Matas, J. 1998, ITPAM, 20, 226
Kosovich, P. A., Kosovichev, A. G., Sadykov, V. M., et al. 2024, ApJ, 972, 169
Laurenza, M., Cliver, E., Hewitt, J., et al. 2009, SpWea, 7, S04008
Laurenza, M., Stumpo, M., Zucca, P., et al. 2024, JSWSC, 14, 8
Lavasa, E., Giannopoulos, G., Papaioannou, A., et al. 2021, SoPh, 296, 107
Le, G.-M., & Zhang, X.-F. 2017, RAA, 17, 123
Liaw, A. 2002, R News, 2/3, 18
Lines, J., Taylor, S., & Bagnall, A. 2016, in 2016 IEEE 16th Int. Conf. on Data Mining (ICDM) (Piscataway, NJ: IEEE), 1041
Lowe, D. G. 2004, Int. J. Comput. Vis., 60, 91
Mazur, J., Mason, G., & Mewaldt, R. 2002, ApJ, 566, 555
Mewaldt, R. 2013, SSRv, 176, 365
Nedal, M., Kozarev, K., Arsenov, N., & Zhang, P. 2023, JSWSC, 13, 26
Núñez, M. 2011, SpWea, 9, S07003
Núñez, M., & Paul-Pena, D. 2020, Univ, 6, 161
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. 2018, Advances in Neural Information Processing Systems 31, ed. S. Bengio et al. (NeurIPS), https://papers.nips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

Reames, D. V. 1999, SSRv, 90, 413

Reames, D. V. 2013, SSRv, 175, 53

Richardson, I., Von Rosenvinge, T., Cane, H., et al. 2014, SoPh, 289, 3059

Richardson, I., & Cane, H. 2010, SoPh, 264, 189

Rotti, S., Aydin, B., Georgoulis, M., & Martens, P. 2022, GSEP Dataset, v5, Harvard Dataverse, doi:10.7910/DVN/DZYLHK

Rotti, S., & Martens, P. C. 2024, ApJS, 267, 40

Rotti, S. A., Aydin, B., & Martens, P. C. 2024, ApJ, 974, 188

Sato, T., Kataoka, R., Shiota, D., et al. 2018, SpWea, 16, 924

Schmidhuber, J. 2015, NN, 61, 85

Schölkopf, B. 2002, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Cambridge, MA: MIT Press)

Schrijver, C. J., & Siscoe, G. L. 2010, Heliophysics: Space Storms and Radiation: Causes and Effects (Cambridge: Cambridge Univ. Press)

Smart, D., & Shea, M. 2003, AdSpR, 32, 109

Stumpo, M., Benella, S., Laurenza, M., et al. 2021, SpWea, 19, e2021SW002794

Suzuki, S., & Abe, K. 1985, Comput. Vis. Graph. Image Process., 30, 32

Tsurutani, B. T., Gonzalez, W., Lakhina, G., & Alex, S. 2003, JGRA, 108, 1268

Van der Maaten, L., & Hinton, G. 2008, JMLR, 9, 2579

Vaswani, A. 2017, Advances in Neural Information Processing Systems 30, ed. I. Guyon et al. (NeurIPS), https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html