Explainable AI for Operational Space Weather Forecasting: An Interpretable Prototype for the EMBRACE-INPE Monitoring Network

Reinaldo R. Rosa[1,*], Andressa Ferreira[1], Salatiel Jordão[1], Carlos E. Falandes[1], Mariana Rubet[1], André R.F. Martinon[1], Stephan Stephany[1], Francisco C. R. Fernandes[1] and Juan A. Valdivia[1]

[1]*Laboratório de Aprendizagem de Máquina Interpretável em Física de Eventos Extremos (LAMIFE), Coordenadoria de Pesquisa Aplicada e Desenvolvimento Tecnológico (COPDT), Instituto Nacional de Pesquisas Espaciais (INPE-MCTI), São José dos Campos, Brazil*

Abstract

Operational space weather monitoring requires not only accurate forecasts but also transparent and interpretable decision-support tools. We present an Explainable Artificial Intelligence (XAI) prototype designed for the EMBRACE-INPE environment. The system integrates GOES X-ray flux, SYM-H index, and AE index, capturing the causal propagation from solar activity to geomagnetic and auroral responses. A sequence-based forecasting module produces six-hour predictions with hyperparameters optimized via Optuna. The XAI layer incorporates uncertainty bands, lag-aware propagation markers, and interpretable multivariate relationships. A three-dimensional regression representation exposes nonlinear dependencies between delayed solar drivers and auroral activity. The prototype is implemented as an interactive dashboard supporting human forecasters. Results illustrate how explainable AI can be embedded directly into operational interfaces, bridging machine learning forecasts and human-centered situational awareness.

Keywords

Explainable Artificial Intelligence, Space Weather, Operational AI, Causality, Time Series Forecasting

1. Introduction

Space weather forecasting requires the continuous integration of heterogeneous solar and geophysical observations into actionable operational information. Modern monitoring systems combine remote sensing data, radiometric measurements, and derived physical indices to anticipate disturbances that may affect satellites, communication systems, navigation infrastructure, and power grids. In this context, machine learning has emerged as a powerful tool for short-term prediction, particularly in problems involving nonlinear coupling and delayed responses across multiple physical scales. Despite these advances, predictive performance alone is insufficient for operational deployment. When artificial intelligence is introduced into real-time monitoring environments, its outputs must be interpretable, traceable, and compatible with human decision-making processes. Forecasters must be able to assess not only what is being predicted, but also how the prediction was produced, which inputs were most influential, and whether the inferred relationships are physically plausible. In this context, Explainable Artificial Intelligence (XAI) provides a framework to address this challenge [1, 2], while recent studies highlight the growing role of machine learning in space weather forecasting [3, 4].

Late-breaking work, Demos and Doctoral Consortium, colocated with the 4th World Conference on eXplainable Artificial Intelligence: July 01–03, 2026, Fortaleza, Brazil

[*]Corresponding author.
reinaldo.rosa@inpe.br (R. R. Rosa); andressa.ferreira@inpe.br (A. Ferreira); salatiel.jordão@inpe.br (S. Jordão); carlos.falandes@inpe.br (C. E. Falandes); rubetmariana@gmail.com (M. Rubet); andre.martinon@inpe.br (A. R.F. Martinon); stephan.stephany@inpe.br (S. Stephany); francisco.fernandes@inpe.br (F. C. R. Fernandes); alejo@uchile.cl (J. A. Valdivia)
0000-0002-2962-4322 (R. R. Rosa); 0009-0009-1001-6798 (A. Ferreira); 0000-0003-3226-9955 (C. E. Falandes); 0000-0001-9439-3121 (M. Rubet); 0000-0002-6578-0170 (A. R.F. Martinon); 0000-0002-6302-4259 (S. Stephany); 0000-0003-2922-3358 (F. C. R. Fernandes); 0000-0003-3381-9904 (J. A. Valdivia)

Recent advances in explainable AI have emphasized the importance of interpretable models in high-stakes environments [2]. In time series analysis, XAI methods have been increasingly developed to capture temporal dependencies and provide meaningful explanations of model behavior [5]. In addition, recent studies have explored the integration of explainability with uncertainty quantification, highlighting the importance of reliable and interpretable predictions in complex systems [6]. In the context of space weather, integrating explainability into operational systems remains an open challenge, particularly when dealing with multiscale and delayed physical processes.

Previous editions of the XAI conference have presented demo-oriented systems in domains such as healthcare, finance, and computer vision. In contrast, the present work focuses on operational space weather forecasting, where explainability must be integrated with physically grounded models and real-time decision support. This combination of causal modeling, temporal forecasting, and human-centered visualization distinguishes the proposed system from existing demonstrations.

In this work, we present a prototype designed for the EMBRACE-INPE [7] monitoring environment, emphasizing the integration of forecasting and explainability into a unified framework. The central premise is that explainability should not be treated as a post hoc diagnostic, but rather as a structural component of the forecasting system itself. Figure 1 illustrates the overall architecture of the proposed pipeline, highlighting the interaction between data ingestion, predictive modeling, and explainability layers. In operational settings such as EMBRACE-INPE, AI systems must support human decision-making rather than act as opaque predictors. In this context, the proposed XAI pipeline integrates forecasting, causal structure, and explainability into a unified operational framework where interpretability becomes a fundamental design principle. In contrast to some previously reported XAI demonstrations, which often focus on static datasets or domain-agnostic benchmarks, the present work emphasizes real-time operational forecasting in a physically grounded environment. The proposed system integrates causal temporal modeling, multivariate dependencies, and human-centered visualization within a unified framework, distinguishing it from prior demo-oriented contributions presented in earlier XAI conference editions.

2. Causal Structure and Forecasting Approach

The forecasting problem is formulated based on three representative signals capturing the Sun–Earth interaction chain: the GOES X-ray flux as a proxy for solar radiative activity, the SYM-H index as a measure of magnetospheric disturbance, and the AE index as an indicator of auroral electrojet activity. These variables are intrinsically coupled through physical processes that propagate energy and disturbances from the Sun to the Earth's near-space environment [8, 9, 10, 11]. This coupling introduces characteristic temporal delays. Solar flares observed in GOES X-ray flux typically precede geomagnetic disturbances represented by SYM-H on timescales ranging from several hours to about one day, depending on propagation conditions. In turn, auroral responses measured by AE follow magnetospheric variations with shorter delays, typically of the order of tens of minutes to a few hours. These characteristic delays define a causal structure that can be exploited for predictive modeling while remaining adaptable within a data-driven framework. Thus, this relationship can be expressed in a generalized form as:

$$AE(t) \approx f\big(\text{GOES}(t-\tau_1),\ \text{SYM-H}(t-\tau_2)\big) \tag{1}$$

where τ_1 and τ_2 represent characteristic time delays associated with the propagation of disturbances from solar activity to magnetospheric and auroral responses. These delays are not fixed a priori but can be inferred from data or implicitly learned by the sequence model. This formulation generalizes the causal structure while preserving its physical interpretation. Therefore, to capture these temporal and nonlinear relationships, the system employs a sequence-based model using Long Short-Term Memory (LSTM) networks. The LSTM model receives as input

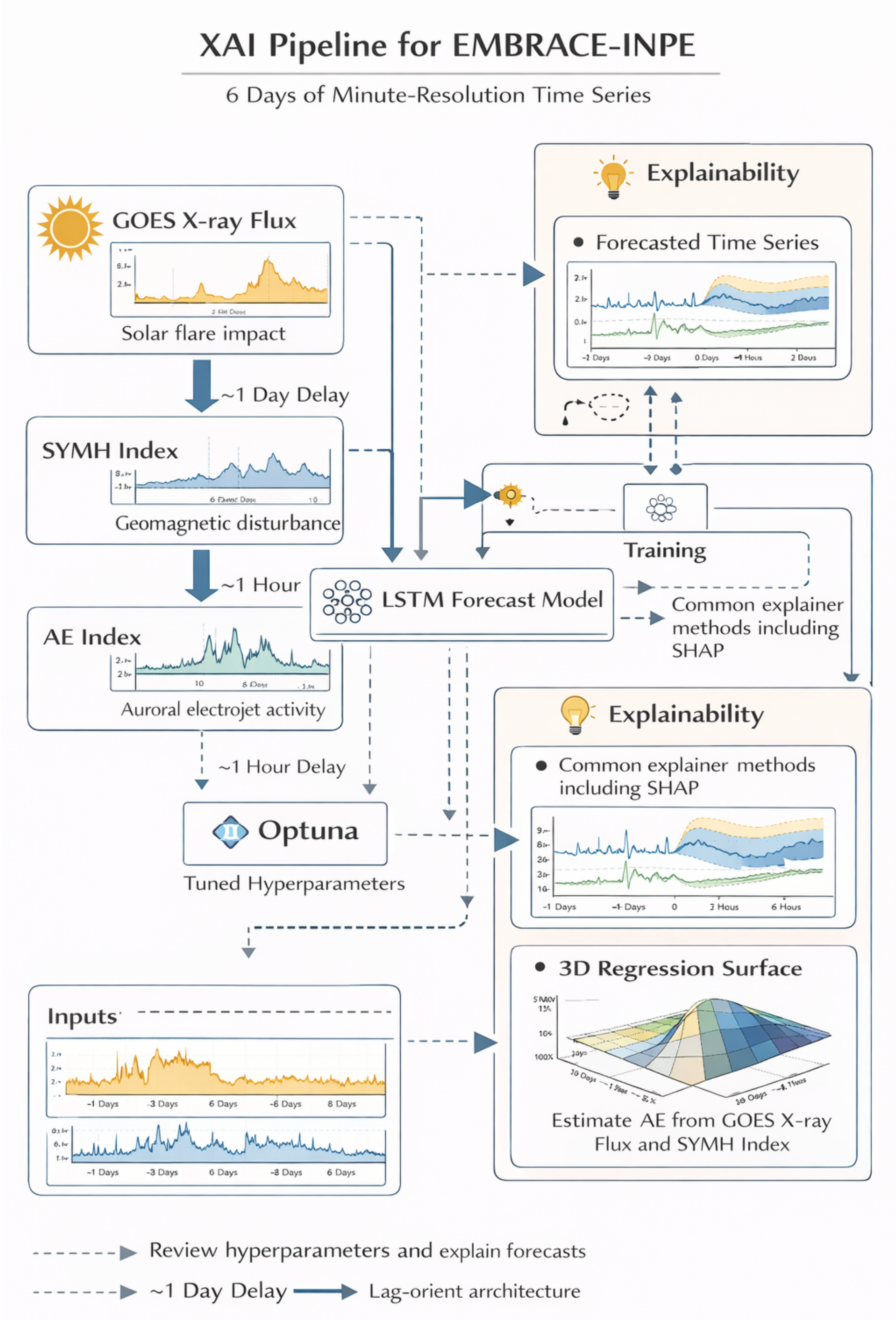

Figure 1: XAI pipeline for EMBRACE-INPE integrating GOES, SYM-H, and AE time series, LSTM forecasting, Optuna optimization, and explainability layers.

a multivariate time series composed of GOES X-ray flux, SYM-H, and AE indices, sampled at one-minute resolution. Input sequences are constructed using sliding windows of length T, allowing the model to capture both short-term fluctuations and delayed responses across multiple temporal scales.

The network architecture consists of one or more stacked LSTM layers followed by fully connected layers that map latent representations to forecasted AE values. The model is trained to produce short-term forecasts with a six-hour horizon, capturing both immediate dynamics and delayed effects. Rather than explicitly fixing the propagation delays, the LSTM structure leverages its internal memory to learn temporal dependencies directly from the data. The training process minimizes a regression loss function, such as the mean squared error, between predicted and observed AE values over the forecasting horizon. Hyperparameters including sequence length, number of hidden units, learning rate, and batch size are optimized using the Optuna framework through Bayesian search. This approach enables efficient exploration of the parameter space while maintaining robustness under varying geomagnetic activity conditions. In practice, the temporal window used for training may span several days of observations, allowing the model to incorporate long-range dependencies while producing short-term forecasts. This configuration provides flexibility to adapt the effective temporal delays according to the data, reinforcing the consistency between machine learning predictions and the underlying physical processes.

3. XAI-Driven Operational Dashboard

A central component of the proposed system is the integration of explainability directly into the operational interface. Rather than treating model interpretation as a separate analytical step, the prototype embeds XAI elements into a dashboard designed to support real-time monitoring and decision-making within the EMBRACE-INPE environment. This design follows a multi-layer architecture in which data representation, predictive modeling, and explainability are progressively integrated into a unified operational view.

At the first layer, the system provides direct visualization of observed data streams, including GOES X-ray flux, SYM-H, and AE indices. These time series are presented in a synchronized temporal framework, allowing the user to inspect the evolution of solar activity and its subsequent impact on the magnetosphere and ionosphere. This layer preserves the traditional monitoring paradigm and ensures continuity with existing operational workflows.

The second layer introduces predictive capabilities through the integration of LSTM-based forecasts. Short-term predictions with a six-hour horizon are displayed alongside observed data, enabling direct comparison between current conditions and expected system evolution. Forecast outputs are accompanied by uncertainty bands, which provide a visual indication of prediction confidence and help analysts assess the reliability of model outputs under varying activity regimes.

The third layer incorporates explainability mechanisms that expose the internal behavior of the predictive model. These include lag-aware propagation markers that highlight the temporal relationship between solar drivers and geomagnetic responses, as well as compact visual indicators summarizing recent activity patterns. In this layer, explainability is tightly coupled with domain knowledge, allowing users to interpret forecasts in terms of physically meaningful processes rather than abstract model behavior.

Figure 2 illustrates how these layers are integrated into the EMBRACE-INPE operational environment. The dashboard combines observed signals, predictive outputs, and explainability cues in a coherent visual structure that supports rapid situational awareness. By aligning machine learning outputs with the cognitive workflow of human forecasters, the system enables a transition from descriptive monitoring to explainable predictive analytics.

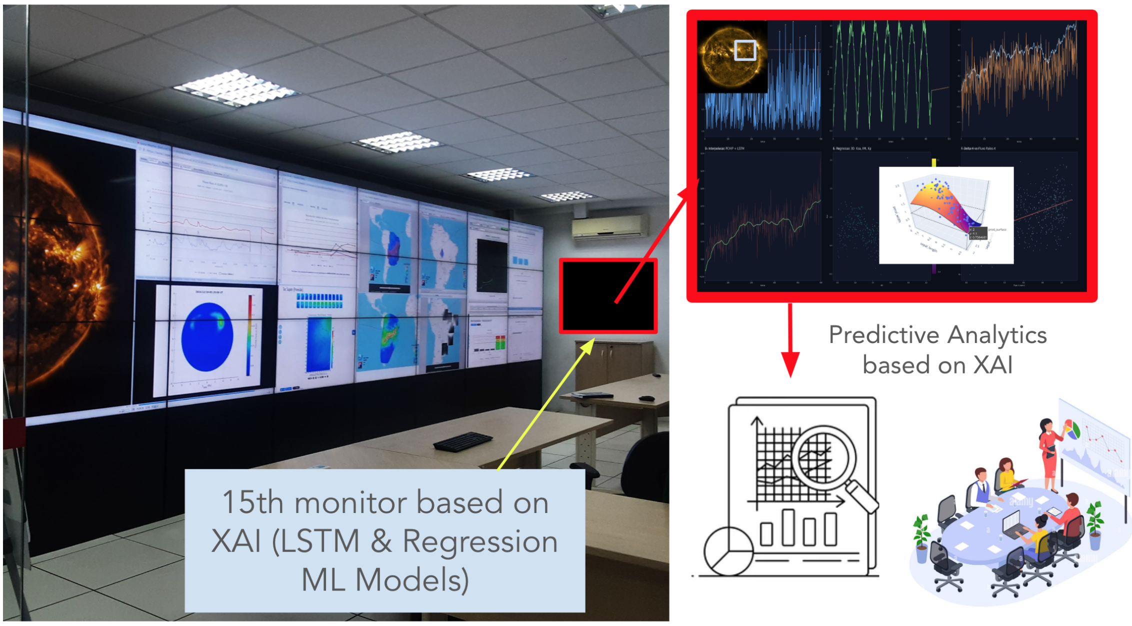

Figure 2: Integration of the XAI dashboard into the EMBRACE-INPE operational environment, illustrating the multi-layer structure combining observational data, LSTM forecasts, and explainability cues.

This multi-layer approach reinforces the idea that explainability is not an isolated component, but an intrinsic property of the operational system. By embedding interpretability into each stage of the visualization pipeline, the dashboard transforms machine learning predictions into actionable and trustworthy information for real-time decision-making.

4. Multivariate Explainability and Model Interpretation

To complement time-series forecasting, the system incorporates a multilevel explainability framework designed to expose the relationships between input variables and predicted outputs in an interpretable and physically meaningful way. Unlike traditional post hoc approaches, explainability is treated here as an intrinsic component of the modeling pipeline, bridging predictive performance and domain understanding. At the temporal level, interpretability is achieved through lag-aware analysis of the input signals. By incorporating delayed versions of GOES X-ray flux and SYM-H into the modeling process, parameterized through characteristic time lags τ_1 and τ_2, the system enables the identification of causal propagation patterns linking solar activity to magnetospheric and auroral responses. These delays are not fixed but are inferred from data or implicitly learned by the sequence model. This approach is consistent with recent developments in time series explainability, where capturing temporal dependencies is essential for meaningful interpretation [5]. The model output can therefore be interpreted in terms of physically grounded delays, rather than purely statistical correlations.

At the model level, interpretability is enhanced through the inspection of multivariate relationships learned during training. To this end, we construct a three-dimensional regression representation that maps delayed GOES and SYM-H inputs to AE predictions. This representation acts as a surrogate model that approximates the learned nonlinear function $f(\cdot)$, providing a global view of how variations in upstream drivers influence auroral activity. Figure 3 illustrates this relationship, revealing structured dependencies that are not directly observable in one-dimensional time series.

At the uncertainty level, the system incorporates confidence-aware representations that allow users to assess the reliability of predictions. Forecast uncertainty is visualized through prediction

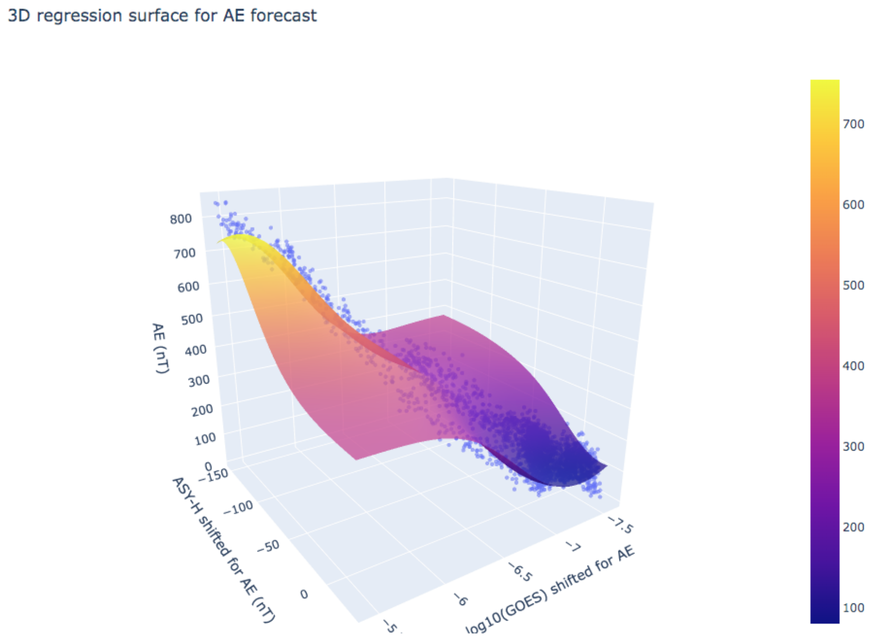

Figure 3: Three-dimensional regression surface relating delayed GOES X-ray flux and SYM-H index to AE activity, providing an interpretable approximation of the learned nonlinear mapping.

intervals, enabling analysts to distinguish between stable and highly variable regimes. This integration of explainability and uncertainty quantification follows recent efforts to combine interpretability with reliability assessment in complex systems [6].

Finally, at the operational level, explainability is embedded into the visualization interface, where model outputs are presented together with contextual information and compact explanatory cues. These include temporal markers indicating propagation delays, visual summaries of recent activity, and interpretable mappings between input space and predicted behavior. This design aligns with the concept of human-centered XAI, in which explanations are tailored to support real-time decision-making rather than purely technical inspection.

By integrating temporal, multivariate, and uncertainty-aware explanations, the proposed framework transforms the model into both a predictive and an interpretable system. This multi-level approach ensures that machine learning outputs remain consistent with domain knowledge while providing actionable insights for operational space weather forecasting.

5. Discussion and Conclusion

The proposed system illustrates a transition from performance-driven machine learning toward explainable and operationally meaningful artificial intelligence in space weather forecasting. By embedding physically motivated temporal delays between solar activity, magnetospheric response, and auroral dynamics, the model incorporates a causal structure directly into the learning process. This approach constrains predictions to domain-consistent relationships while reducing the risk of spurious correlations. In addition, the integration of explainability into the dashboard interface enables forecasts to be interpreted in the context of observed signals, uncertainty levels, and propagation patterns, thereby supporting real-time situational awareness in the EMBRACE-INPE operational environment. The multi-layer design further reinforces that explainability is not an auxiliary feature, but a structural component of the forecasting system.

From a broader perspective, this work highlights the importance of treating explainability as a core design principle in scientific and operational AI systems. By combining sequence-based modeling, multivariate interpretation, and human-centered visualization, the proposed framework bridges the gap between machine learning outputs and physically interpretable processes. The integration of uncertainty-aware representations further strengthens the reliability of the system, enabling users to assess prediction confidence in dynamic conditions. Future work will focus on the incorporation of real observational datasets, systematic evaluation of XAI techniques for time series forecasting, and validation with operational forecasters. Therefore, the proposed XAI framework is not restricted to space weather applications and can be adapted to other domains involving the monitoring and forecasting of complex and extreme events. In particular, similar approaches can be customized for meteorological forecasting and natural disaster risk assessment, where multiscale interactions, temporal delays, and uncertainty play a critical role. In this context, institutions such as CEMADEN may benefit from XAI-driven systems that integrate predictive modeling with interpretable outputs, supporting decision-making processes in scenarios involving extreme environmental events. These developments are expected to advance the role of explainable AI as a trustworthy component of decision-support systems in space weather and other critical scientific domains.

Acknowledgments

The authors acknowledge INPE-MCTI and the EMBRACE initiative for providing the scientific and operational context that motivated this work. The authors also acknowledge the activities of LAM-IFE/COPDT in the development of interpretable AI approaches for critical physical systems. This work was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under Grant 446053/2023-6, and FAPESP (Proc. No 21.15114-8). R.R.R. acknowledges fruitful discussions on EMBRACE data and AI-based approaches with INPE researchers Alisson Del Lago, Joaquim Costa, José R. Cecatto, Laysa Resende, Luis E. Vieira, Siomel Savio, Valdivino Santiago Junior and Walter Gonzalez. The importance of XAI in the context addressed in this work has also been discussed within CEMADEN with Dr. Leonardo B. L. Santos.

Declaration on Generative AI

During the preparation of this work, the authors used generative AI tools, based in very consistent lexicon construction, for grammar checking (Claude and ChatGPT), text refinement (Claude), structuring support and drafting assistance (Figure 1) (ChatGPT and Julius). All generated content was carefully reviewed and edited by the authors, who take full responsibility for the publication's content.

References

[1] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608 (2017).

[2] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (2019) 206–215. doi:10.1038/s42256-019-0048-x.

[3] E. Camporeale, The challenge of machine learning in space weather: Nowcasting and forecasting, Space Weather 17 (2019) 1166–1207. doi:10.1029/2018SW002061.

[4] M. G. Bobra, S. Couvidat, Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm, The Astrophysical Journal 798 (2015) 135. doi:10.1088/0004-637X/798/2/135.

[5] E. T. Mekonnen, Explaining time series classifiers through post-hoc xai methods capturing temporal dependencies, in: Proceedings of the 4th World Conference on Explainable Artificial Intelligence (XAI 2026), 2026. URL: https://ceur-ws.org/Vol-4017/paper_42.pdf.

[6] E. Schiller, Explaining uncertainty: Exploring the synergies of explainable artificial intelligence and uncertainty quantification, in: Proceedings of the 4th World Conference on Explainable Artificial Intelligence (XAI 2026), 2026. URL: https://ceur-ws.org/Vol-4017/paper_47.pdf.

[7] EMBRACE/INPE, Estudo e monitoramento brasileiro do clima espacial, 2026. Operational monitoring initiative of INPE.

[8] C. R. Neto, A. Zanandrea, F. M. Ramos, R. R. Rosa, M. J. A. Bolzan, L. D. A. Sá, Multiscale analysis from turbulent time series with wavelet transform, Physica A 295 (2001) 215–218. doi:10.1016/S0378-4371(01)00076-0.

[9] R. R. Rosa, M. Karlický, T. B. Veronese, N. L. Vijaykumar, H. S. Sawant, A. I. Borgazzi, M. S. Dantas, E. B. M. Barbosa, R. A. Sych, O. Mendes, Gradient pattern analysis of short solar radio bursts, Advances in Space Research 42 (2008) 844–851. doi:10.1016/j.asr.2007.08.015.

[10] T. B. Veronese, R. R. Rosa, M. J. A. Bolzan, F. C. R. Fernandes, H. S. Sawant, M. Karlický, Fluctuation analysis of solar radio bursts associated with geoeffective x-class flares, Journal of Atmospheric and Solar-Terrestrial Physics 73 (2011) 1311–1316.

[11] M. J. A. Bolzan, R. R. Rosa, Multifractal analysis of interplanetary magnetic field obtained during cme events, Annales Geophysicae 30 (2012) 1107–1112. doi:10.5194/angeo-30-1107-2012.