POLITECNICO DI TORINO
Repository ISTITUZIONALE

Forecasting Geoffective Events from Solar Wind Data and Evaluating the Most Predictive Features through Machine Learning Approaches

(Article begins on next page)

26 December 2024

Forecasting Geoffective Events from Solar Wind Data and Evaluating the Most Predictive Features through Machine Learning Approaches

Sabrina Guastavino[1,2] , Katsiaryna Bahamazava[3] , Emma Perracchione[3] , Fabiana Camattari[1,2], Gianluca Audone[3] ,
Daniele Telloni[2] , Roberto Susino[2] , Gianalfredo Nicolini[2] , Silvano Fineschi[2], Michele Piana[1,2] , and
Anna Maria Massone[1]

[1] MIDA, Dipartimento di Matematica, Universitá di Genova, via Dodecaneso 35 16146 Genova, Italy; guastavino@dima.unige.it
[2] Istituto Nazionale di Astrofisica (INAF), Osservatorio Astrofisico di Torino, Italy
[3] Dipartimento di Scienze Matematiche Giuseppe Luigi Lagrange, Politecnico di Torino, Torino, Italy

Abstract

This study addresses the prediction of geomagnetic disturbances by exploiting machine learning techniques. Specifically, the Long Short-term Memory recurrent neural network, which is particularly suited for application over long time series, is employed in the analysis of in situ measurements of solar wind plasma and magnetic field acquired over more than one solar cycle, from 2005 to 2019, at the Lagrangian point L1. The problem is approached as a binary classification aiming to predict 1 hr in advance a decrease in the SYM-H geomagnetic activity index below the threshold of -50 nT, which is generally regarded as indicative of magnetospheric perturbations. The strong class imbalance issue is tackled by using an appropriate loss function tailored to optimize appropriate skill scores in the training phase of the neural network. Beside classical skill scores, value-weighted skill scores are then employed to evaluate predictions, suitable in the study of problems, such as the one faced here, characterized by strong temporal variability. For the first time, the content of magnetic helicity and energy carried by solar transients, associated with their detection and likelihood of geoeffectiveness, were considered as input features of the network architecture. Their predictive capabilities are demonstrated through a correlation-driven feature selection method to rank the most relevant characteristics involved in the neural network prediction model. The optimal performance of the adopted neural network in properly forecasting the onset of geomagnetic storms, which is a crucial point for giving real warnings in an operational setting, is finally showed.

Unified Astronomy Thesaurus concepts: Space weather (2037); Solar wind (1534); Neural networks (1933); Classification (1907)

1. Introduction

Space weather (SW) science studies how solar–terrestrial interactions affect the geospace environment (Pulkkinen 2007); specifically, it involves predicting major solar disturbances that can pose significant risks to terrestrial facilities, with potential economic and security implications. The SW domain thus covers the various physical processes involved in the transfer of energy from solar wind and events to the Earth system, such as magnetic reconnection, the generation of ring currents in the terrestrial magnetosphere, and the interaction of solar charged particles with the Earth's atmosphere (Chapman & Ferraro 1931; Dungey 1961; Fairfield & Cahill 1966). The most energetic events on the solar surface are flares (Piana et al. 2022), intense electromagnetic emissions that can accelerate particles to relativistic velocities. They are often followed by massive eruptive events of chromospheric and coronal material, known as coronal mass ejections (CMEs; Webb & Howard 2012), during which the solar plasma and the magnetic field are expelled into interplanetary space. Flares and CMEs, along with high-speed streams (HSSs) of particles, and corotating interaction regions and heliospheric current sheet crossings, can initiate geomagnetic storms that, when particularly intense, can have severe consequences for human activities and ground- and space-based infrastructures, such as telecommunications systems, satellite orbits, or power grids and pipelines.

Geomagnetic disturbances are usually detected by measuring perturbations in magnetospheric electric currents induced by solar storms. Specifically, the Disturbance Storm Time (DST; amplitude planetary, Ap) geomagnetic index, derived from a network of ground-based magnetometer stations near the equator (at subauroral latitudes), is an estimate of the variations in the horizontal component of the ring currents circling the Earth in the equatorial plane (of the electric currents aligned with the Earth's magnetic field in the auroral ionosphere). DST and Ap thus provide an assessment of the severity of geomagnetic storms at low and high latitudes, respectively. The DST index (acquired hourly) or the equivalent SYM-H (acquired every minute) are those most commonly deployed to reveal geomagnetic storms and measure their degree of geoeffectiveness. These are classified as moderate, intense, or extreme, depending on whether the DST index falls below thresholds of -50, -100, and -250 nT (Cander & Mihajlovic 1998).

Recently, machine learning and deep learning techniques are increasingly being employed in SW applications (Camporeale et al. 2018), especially in the prediction of solar flares (Bobra & Couvidat 2015; Barnes et al. 2016; Campi et al. 2019; Georgoulis et al. 2021, 2024; Guastavino et al. 2022a, 2023a), the onset of CMEs, and their arrival time on Earth (Guastavino et al. 2023b; Singh et al. 2023). Such studies are based on remote observations of the Sun and its atmosphere and, more specifically, on the identification of CMEs and the

extraction of their morphological/dynamic properties (such as angular width, velocity, and acceleration) from time sequences of coronographic white-light images, such as those from the Large Angle Spectroscopic Coronograph, as in Pricopi et al. (2022) and Vourlidas et al. (2019).

Less attention has been devoted, however, to predicting the degree of geoeffectiveness of solar events impinging on the Earth. In Telloni et al. (2023), the problem was addressed using interplanetary plasma and magnetic field measurements (and, in particular, the intensity and the normal component of the ecliptic plane of the magnetic field vector, and the bulk speed, temperature, and density of the solar wind plasma) acquired in situ at the L1 Lagrangian point, 1.5 million km from Earth in the sunward direction, by resorting to the use of different types of neural networks. The recurrent architectures were found to be the best in predicting geomagnetic events associated with SYM-H values below -50 nT, achieving 94% (70%) accuracy when the prediction was made 1 (8) hr in advance.

Similarly to Telloni et al. (2023), this paper aims to forecast the degree of severity of CME-induced geomagnetic storms but explores and extends its predictive capabilities, using more refined and potentially better approaches. More specifically, as in the previous work, the problem was tackled as a binary classification with the goal of forecasting 1 hr ahead a decrease in the SYM-H geomagnetic index below -50 nT, using the Long Short-term Memory (LSTM) recurrent network (which was shown to have the best performance in Telloni et al. 2023) applied to long time series of solar wind data. However, unlike Telloni et al. (2023), in which the strong data imbalance (SYM-H < -50 nT only 2% of the time) potentially affecting the proper neural network training was addressed by exploiting data augmentation, in the present study, this issue is faced by using an appropriate loss function designed to automatically optimize a suitable skill score for evaluating predictions in the case of highly imbalanced data sets (Marchetti et al. 2022). The advantage of this approach is to avoid data handling. The prediction performances are also evaluated against value-weighted skill scores (Guastavino et al. 2022b) that are more suited for forecasting over time (Hu et al. 2022; Guastavino et al. 2022c) and will be shown to better evaluate performance in predicting the onset of geomagnetic storms. The method provides scores comparable to those of a simple persistence model. As an important difference, while the latter obviously cannot predict the onset of geomagnetic storms but only the events that occur during the recovery phase, the LSTM model not only forecasts events within the recovery phase of the storm but also its onset with about 98% of accuracy. In addition, we have computed an estimate of the uncertainty of prediction by performing several runs of the neural network with several generations of training, validation, and test sets. More importantly, in the present analysis, besides all interplanetary parameters (hereafter features) directly acquired by the instruments on board the spacecraft, i.e., plasma and magnetic field measurements of the solar wind, some derived quantities, such as magnetic helicity (which has been shown to be critical for properly detecting CMEs at L1, e.g., Telloni et al. 2019), and solar wind transported energy (which has been shown to play a crucial role in the magnetospheric response to solar drivers, Telloni et al. 2020a), were also considered as inputs to the neural network. Finally, a correlation-driven feature selection method is here used to rank the most relevant, i.e.,

predictive, features involved in the neural network prediction model.

The layout of the paper is as follows: Section 2 presents the solar wind data set used as input for the neural network model; Section 3 describes the machine learning approach and correlation-driven feature ranking method; and Section 4 shows the forecasting performances and the most predictive features. Our conclusions are offered in Section 5.

2. Data Set

It is well known that a huge amount of data is required to train, validate, and test neural networks. The data set used in this paper consists of 7,888,319 one minute acquisitions related to multispacecraft interplanetary parameters acquired at L1 and geomagnetic activity indices measured on the ground. The time period spans the years 2005–2019, thus encompassing the entire solar cycle 24 and the descending (ascending) phase of the previous (subsequent) solar cycle 23 (25). The Operating Missions as a Node on the Internet (King & Papitashvili 2005) database was used, which is intended specifically to study the effects of the solar wind variations on the Earth's magnetosphere. In fact, the interplanetary measurements are artificially time shifted forward as if they had been acquired at the same location as the magnetospheric indices (in other words, the distance of 1.5 million km separating L1 from Earth has been nullified by adding to the acquisition time of the heliospheric quantities the time it takes the solar wind to reach Earth). This allows the temporal correlations between solar and magnetospheric indices to be explored directly. Among all available interplanetary measurements, the following parameters were taken into account in the study: the magnitude and components of the magnetic field vector $\boldsymbol{B}$, the components of the flow velocity vector $\boldsymbol{V}$, along with the bulk speed U, density ρ, and temperature T of the solar wind plasma. Some quantities derived from spacecraft measurements were also considered. These are the magnetic helicity H_m, the kinetic E_k and magnetic energy E_m, and the total energy $E = E_k + E_m$. For a comprehensive discussion on how these quantities are estimated in the solar wind, the reader is referred to Telloni et al. (2019, 2020a). Here it is only noted that magnetic helicity is a measure of the degree of twisting of the magnetic field lines and is therefore a clear signature of the CME-embedded flux rope (which is a helical structure and therefore can be seen as a region of high magnetic helicity). It follows that H_m is beneficial for a proper detection of CMEs. On the other hand, it appears evident that the more energetic the CME is (either due to its high velocity and/or the intensity of the associated magnetic field), the more severe is the expected induced geomagnetic event since more energy will be transferred to the geospace. As a result, E is a key parameter for correct prediction of geoeffectiveness of solar storms. As for the assessment of geomagnetic activity, the SYM-H index was instead employed.

As mentioned above, the goal of this study is to predict whether, based on a history of features over the past 24 hr, in the next hour the SYM-H geomagnetic index will drop below the -50 nT threshold, customarily referred to for potential severe magnetospheric disturbances. From a purely computational perspective (i.e., to save machine resources), the time series are resampled at a 1 hr resolution. This means that the total number of samples is thus reduced to just over 130,000. Since the study is approached as a binary prediction problem,

samples corresponding to time periods when SYM-H $\leqslant -50\,$nT are labeled with 1 or 0, respectively, and named 1- or 0-labeled samples. As already noted in Telloni et al. (2023) and reported above, the data set is highly imbalanced: only 2.53% of total samples are labeled with 1. Section 3 will present a suitable approach associated with the employment of an appropriate loss function to address this problem and enable proper training and validation of the neural network based on the use of suitable metrics for performance evaluation.

3. Prediction and Feature Ranking

3.1. Assessment of Results

In order to compare the performances of machine learning methods for binary classification problems for the prediction of geoeffective events, the following points should be accounted for. First, the classification results should be evaluated by considering skill scores that are suitable for imbalanced data classification. Indeed, the percentage of geoeffective events is really small, as already pointed out in Section 2. Therefore, a chosen score needs to be capable of representing the performance of the classifier concerning the small positive class. The classical skill scores are computed on the entries of the so-called confusion matrix (CM), which is defined as

$$CM = \begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}, \qquad (1)$$

where the four entries are true positives (TPs), i.e., the number of samples labeled with 1 and correctly predicted as positive; true negatives (TNs), i.e., the number of samples labeled with 0 and correctly predicted as negative; false positives (FPs), i.e., the number of samples labeled with 0 incorrectly predicted as positive; and false negatives (FNs), i.e., the number of samples labeled with 1 and incorrectly predicted as negative. Among all possible skill scores, the True Skill Statistic (TSS; Hanssen & Kuipers 1965) is less sensitive to the class-imbalance ratio (Bloomfield et al. 2012) than others, and therefore, it is particularly suitable for evaluation of imbalanced classification tasks. It is defined as the balance between the true positive and true negative rates (named also sensitivity and specificity, respectively), as follows:

$$TSS(CM) = \frac{TP}{TP + FN} + \frac{TN}{FP + TN} - 1 = SENS(CM) + SPEC(CM) - 1 . \qquad (2)$$

Its values range in the interval $[-1, 1]$, and the performance is optimal when TSS = 1, while TSS > 0 means that the rates of positive and negative events are mixed up.

Second, the evaluation of binary predictions performed over time should take into account the forecast value, measured in terms of its usefulness to an operational setting to support the user while making a decision, as the importance in predicting the starting time of a geomagnetic storm. Value-weighted skill scores introduced by Guastavino et al. (2022b) that better account for the intrinsic dynamical nature of forecasting problems are defined on a value-weighted CM that assigns different weights to FPs (denoted as wFPs) and FNs (denoted as wFNs) in such a way as to account for the distribution of predictions along time with respect to the actual occurrences.

By denoting the value-weighted CM as

$$wCM = \begin{pmatrix} TN & wFP \\ wFN & TP \end{pmatrix}, \qquad (3)$$

predictions are assessed by computing the value-weighted TSS (wTSS) defined as follows:

$$wTSS = TSS(wCM). \qquad (4)$$

Finally, the splitting strategy between training, validation, and test sets should take into account the rare-event nature of the problem by maintaining uniformly the percentage of 1-labeled samples and the temporal distribution of events in order to not mix temporally samples in training, validation, and test sets. Furthermore, the splitting strategy should be repeated several times in order to achieve some statistical significance. Therefore, many classification tests should be carried out by generating different triples of training, validation and test sets by maintaining the positive percentage of samples and not drawing samples completely randomly between training and test with respect to time.

3.2. Recurrent Neural Network

As neural network, we adopted the LSTM network (Hochreiter & Schmidhuber 1997), which is the most widely adopted type of recurrent neural network, able to process sequential data by solving the well-known short-term memory problem of basic recurrent architectures. In the experiments, we set the LSTM units equal to 72 followed by a dense layer of 64 neurons where the rectified linear unit activation function is adopted and a final dense sigmoid unit drives the output of the network to be in the interval [0, 1] in order to perform binary prediction. The LSTM network is trained for 100 epochs by taking a batch size of 256 samples; the Adam optimizer (Kingma & Ba 2015) is adopted for the optimization process with learning rate equal to 10^{-4}.

In order to face the class-imbalance issue, we adopted a suitable loss function that is designed to optimize an appropriate score. This strategy is introduced in Marchetti et al. (2022), and it is based on a probabilistic definition $\overline{CM}$ of the classical CM, depending on a chosen cumulative density function (cdf) on [0, 1] for the threshold parameter τ, which separates 0 and 1. Let $(x_i, y_i)_{i=1}^n$ be input-label samples, where $y_i \in \{0, 1\}$ is the actual label associated to the sample x_i, and let $f(x_i) \in (0, 1)$ be the probability outcome of the neural network f on the sample x_i. We defined the probabilistic confusion matrix $\overline{CM}$ as

$$\overline{CM} = \begin{pmatrix} \overline{TN} & \overline{FP} \\ \overline{FN} & \overline{TP} \end{pmatrix}, \qquad (5)$$

characterized by the following entries:

$$\overline{TP} = \sum_{i=1}^n y_i f(x_i), \quad \overline{TN} = \sum_{i=1}^n (1 - y_i)(1 - f(x_i))$$
$$\overline{FP} = \sum_{i=1}^n (1 - y_i)f(x_i), \quad \overline{FN} = \sum_{i=1}^n y_i(1 - f(x_i)). \qquad (6)$$

We chose the score-oriented loss function based on the TSS, which is defined as follows:

$$\ell_{TSS} := -TSS(\overline{CM}). \qquad (7)$$

The main advantage consists of an automatic optimization of the desired skill score during the training phase without the need of a posteriori optimization of the threshold.

3.3. Correlation-driven Feature Ranking

Feature ranking methods are commonly employed to identify a reduced subset of highly predictive features and to assess the relevance of physical attributes. One common technique consists in evaluating how each feature impacts the predictions according to some permutation importance score. Once a model is fitted via a training set, the permutation importance algorithm is implemented to evaluate how the accuracy of the prediction changes when a single feature is randomly shuffled in the validation data set (Breiman 2001). Indeed, when a feature is shuffled, its relevance increases as the accuracy score on the prediction decreases. Nevertheless, the permutation feature importance algorithm does not take into account the correlation between features. Therefore, we consider a correlation-driven feature importance method inspired by Kaneko (2022), which includes the absolute correlation coefficients between features in the permutation process. Hence, in presence of strongly correlated features, such a method leads to more stable and reliable feature ranking schemes. In the following, we summarize the steps of the correlation-driven feature importance algorithm:

1. Train a neural network $\hat{f}$.
2. Compute the prediction of the trained neural network $\hat{f}$ on the validation data, denoted as $Z_{\text{val}} = \{(X_{\text{val}}, y_{\text{val}})\}$ and compute the reference score s_{val}, i.e., the score computed between the prediction $\hat{f}(X_{\text{val}})$ and y_{val}.
3. Calculate the correlation coefficients between all the features by following the procedure in Kaneko (2022); in particular, the correlation coefficient between two features is set to 0 if there is the possibility of a chance correlation.
4. For each feature j, i.e., the jth column of validation data X_{val}, and for each repetition $k \in \{1,\dots,K\}$, randomly shuffle the jth feature of the validation data set X_{val}, and for each feature $l \neq j$ for which the correlation value $c_{j,l}$ is higher than 0, randomly sample the column l of X_{val} (without duplication with a probability of $c_{j,l}$) to generate a corrupted version of the validation data set, denoted as $X_{\text{val}}^{j,k}$.
5. Compute the score $s_{j,k}$ between the prediction on the corrupted data $X_{\text{val}}^{j,k}$, i.e., $\hat{f}(X_{\text{val}}^{j,k})$ and y_{val}.
6. Evaluate the importance $\mathcal{I}_j$ of the jth feature by computing the difference between the reference score and the mean score obtained on the corrupted validation data sets, i.e.,

$$\mathcal{I}_j = s_{\text{val}} - \frac{1}{K}\sum_{k=1}^{K} s_{j,k}. \tag{8}$$

Such a procedure allows obtaining a feature ranking: the higher $\mathcal{I}_j$ is, the higher the contribution of the feature in the prediction.

4. Results

The LSTM network described in Section 3.2 has been applied to time series of features described in Section 2. First, a splitting strategy based on stratified k-fold splitting allows the generation of training, validation, and test set that shares the same rate of about 2.5% of 1-labeled samples, leading to training and validation sets whose sum represents about 75% of the total number of the sample and a test set representing the

Table 1
Feature Rankings with Respect to the Correlation-driven Feature Importance Method Shown in Section 3.3 in the Cases SYM-H Is Included between Features and Not

SYM-H Included		SYM-H Not Included	
Feature	Importance Value	Feature	Importance Value
(1) SYM-H	0.5577	(1) B	0.2936
(2) B	0.3392	(2) Bz	0.2605
(3) Total Energy	0.3334	(3) Total Energy	0.2175
(4) Vx	0.3256	(4) Magnetic Helicity	0.2042
(5) V	0.3162	(5) Vx	0.1902
(6) Temperature	0.287	(6) Magnetic Energy	0.1901
(7) Magnetic Energy	0.2793	(7) Temperature	0.18
(8) Magnetic Helicity	0.2628	(8) V	0.1783
(9) Kinetic Energy	0.2427	(9) Kinetic Energy	0.1604
(10) Bz	0.2303	(10) Density	0.1451
(11) Density	0.2044	(11) Vy	<0.1
(12) Vy	<0.1	(12) By	<0.1
(13) By	<0.1	(13) Bx	<0.1
(14) Vz	<0.1	(14) Vz	<0.1
(15) Bx	<0.1	…	…

remaining 25% of samples. We investigated how the prediction performances change if the SYM-H of the past 24 hr with cadence 1 hr is added as an additional feature to the considered solar wind features, i.e., B, Bx, By, and Bz (which represent the absolute value of the magnetic field vector B and the three components of B); V, Vx, Vy, and Vz (which represent the absolute value of the velocity vector V and the three components of V); ρ (density); T (temperature); E_k (kinetic energy); E_m (magnetic energy); E (total energy); and H_m (magnetic helicity). When the network is trained, we applied the correlation-driven feature importance method described in Section 3.3 in order to rank features according to their predictive capabilities, and we selected the first 10 features that are associated with a meaningful importance value at least higher than 0.1. In Table 1, we report the ranking of features with the associated importance value both in the cases where the SYM-H between features is included and where it is not. In detail, we compute the CM and some common skill scores such as the Heidke skill score (Heidke 1926); the sensitivity (SENS); the specificity (SPEC) defined in Equation (2); the F1 score (F1), which is the harmonic mean of precision and sensitivity; and balanced accuracy (BA), which is the arithmetic mean between sensitivity and specificity. We noticed that the most predictive solar wind features are almost the same, both when SYM-H is included and not between features at least of permutation. In particular, the energies, such as the total energy is the third in both rankings: this is coherent with the analysis conducted between the DST and energy in Telloni et al. (2020b). Further, as we expected, features that are not correlated with the occurrence of geomagnetic storms have a negligible importance value less then 0.1. In Table 2, we report the prediction performances on the test set when all the features were used and if the SYM-H is included in the list of features compared with the performance when only the subset of predictive features is considered. We noticed that the
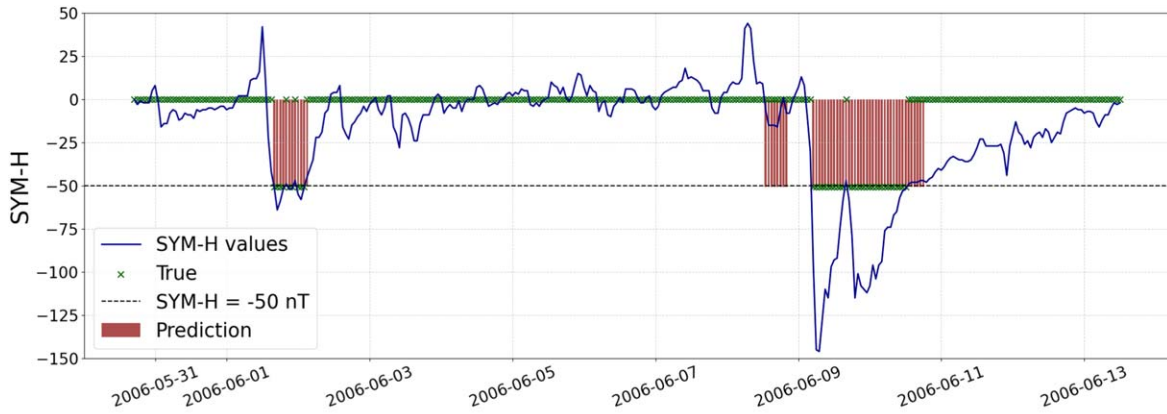
(a) SYM-H not included in features.
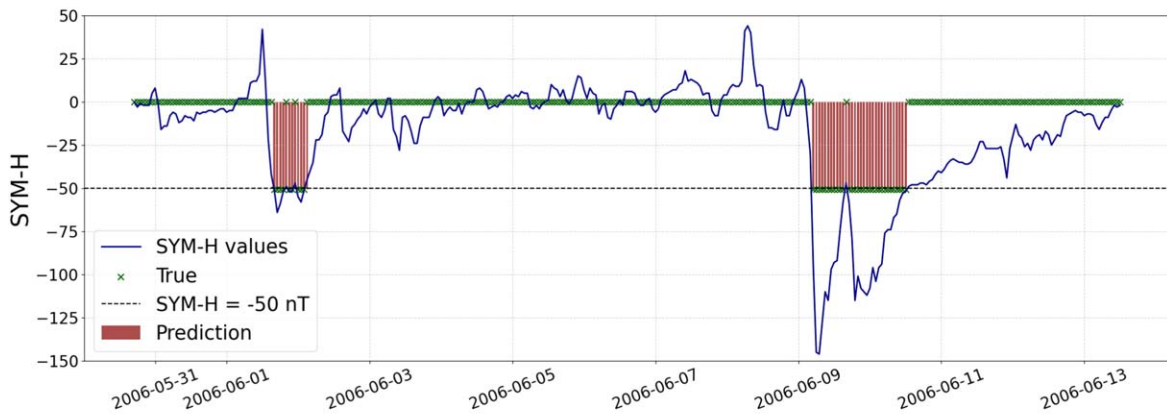
(b) SYM-H included in features.

Figure 1. Predictions over time on a temporal window of the test set of splitting 1 (from 2005 July 24 to 2005 August 2): the top panel represents the prediction when SYM-H is not included in the subset of features, whereas the bottom panel represents the one when SYM-H is included between features.

Table 2
Comparison of the Performances on the Test Set between Using All Features or the Selected Features with the Correlation-driven Feature Importance Method
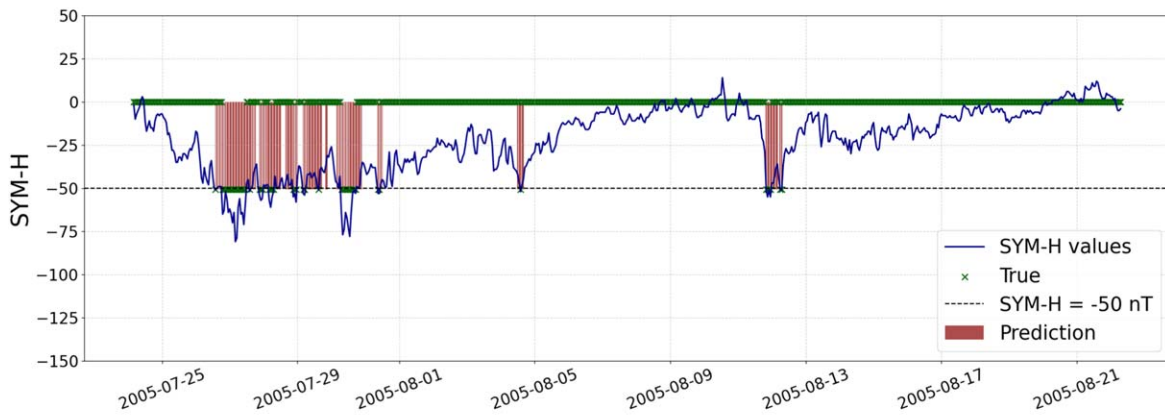
| | SYM-H Included | | | | SYM-H not Included | | | |
	All Features		Selected Features		All Features		Selected Features	
CM	TP = 629 FP = 74	FN = 204 TN = 31599	TP = 707 FP = 430	FN = 124 TN = 615	TP = 495 FP = 36	FN = 336 TN = 668	TP = 550 FP = 430	FN = 281 TN = 31599
TSS	0.742		0.8412		0.5766		0.633	
HSS	0.655		0.7593		0.4963		0.4622	
SENS	0.755		0.8508		0.5957		0.6619	
SPEC	0.987		0.9904		0.9809		0.9715	
F1	0.665		0.766		0.511		0.48	
BA	0.871		0.9206		0.7883		0.8267	

performances are higher when the subset of the most predictive features is adopted: this confirms the importance of a feature selection procedure also in the prediction phase. Finally, in order to assess statistical robustness, we reported results when five different splittings of training, validation, and test sets are considered. In Table 3, we reported the TSS and value-weighted TSS (wTSS) among the five generated test sets. We pointed out that when SYM-H is not used as a feature, the performances are low: this is due to the fact that when a geomagnetic storm is in place, then the SYM-H is below -50 nT, and such information helps the network to learn that if

in the previous hour the SYM-H is below -50 nT, the probability that it is below this threshold in the next hour is high. Further, we made the comparison with a simple persistence model (as in Hu et al. 2023), and the performance in terms of TSS are comparable with the ones obtained by the LSTM model when the SYM-H is included as an additional feature (the persistence model provides a mean TSS on the same test sets about 0.8361 ± 0.028). However, it should be noted that the persistence model can basically predict whether in the next hour the Earth's magnetosphere will still be perturbed (SYM-H < -50 nT), that is, it is able to ascertain the
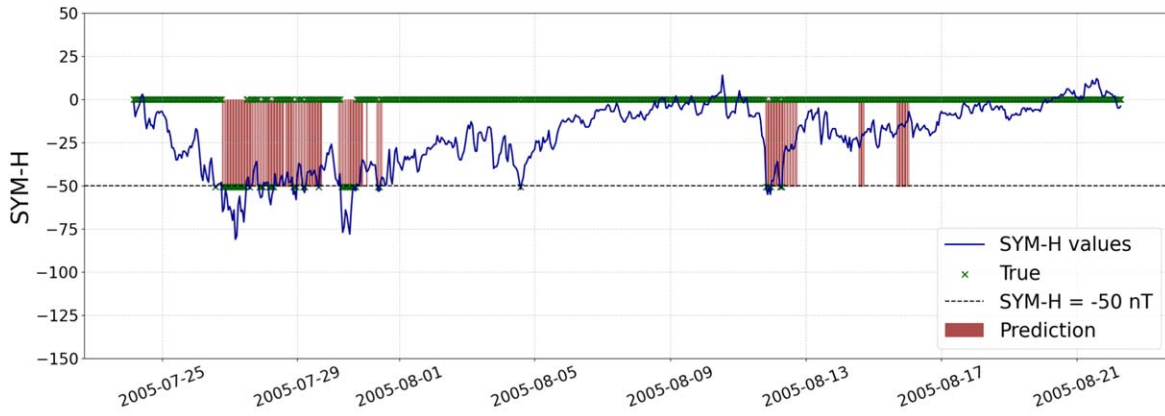
(a) SYM-H not included in features

(b) SYM-H included in features

Figure 2. Predictions over time on a temporal window of the test set of splitting 1 (from 2006 May 30 to 2006 June 12): the top panel represents the prediction when SYM-H is not included in the subset of features, whereas the bottom panel represents the one when SYM-H is included between features.

Table 3
Performances on 5 Test Sets in Case the SYM-H is Included or Not in the Subset of Features

Splitting	TSS		wTSS	
	SYM-H Included	SYM-H not Included	SYM-H Included	SYM-H not Included
1	0.9073	0.82029	0.8964	0.7597
2	0.8654	0.7146	0.8503	0.6166
3	0.9082	0.7791	0.8973	0.6743
4	0.8442	0.6296	0.7838	0.4844
5	0.846	0.6334	0.8164	0.4757
Mean ± std	0.8742 ± 0.028	0.7154 ± 0.076	0.8164 ± 0.044	0.6022 ± 0.109

presence of the recovery phase once the geomagnetic storm has already started. On the other hand, the LSTM model is also able to predict its onset. Specifically, we defined the onset of a geomagnetic disturbance as the first time when SYM-H < -50 nT after a quiet condition (namely, at least for the last 3 hr the SYM-H was above -50 nT). Then, we computed, during the test window of the first splitting, how many times the onset was correctly predicted: it results that 91 over 93 onset events were properly forecasted, i.e., with a 98% accuracy. Therefore, in Figures 1 and 2, we report the distribution of predictions along time in correspondence with two different temporal windows that include some geomagnetic

storms. The period in Figure 1 is between 2005 July 24 and 2005 August 22, and the one in Figure 2 covers from 2006 May 30 to 2006 June 12. Specifically, the red bar plots represent the predicted alarms, and the green crosses represent the actual labels (those placed at 0 nT represent label 0, and those placed at -50 nT represent label 1). In the top panels of Figures 1 and 2, we show the predictions obtained by the LSTM model when SYM-H is not used as an additional feature, whereas in the bottom panels, we show the ones when SYM-H is used as an additional feature. When SYM-H is not used as an additional feature, we noticed that more false positives are present. However, from all panels, it is clear that

the starting time of the geomagnetic storm is well predicted even when SYM-H is not used as a feature: this confirms the importance of solar wind features that give information about the possible occurrence of a geomagnetic storm, whether caused by a CME or other events such as HSSs.

5. Conclusions

In this work, the performance of the LSTM recurrent neural architecture in providing 1 hr advance alerts of critical SYM-H values less than -50 nT and thus indicative of geomagnetic disturbances was estimated. It turns out that its predictive capability reaches a mean TSS over multiple test sets of 0.87 and 0.72 in the case of considering and not considering SYM-H itself, respectively, as an input feature of the neural network. It appears evident that in the former case, the neural network is able to predict the activity status of the Earth's magnetosphere even after the onset of the geomagnetic storm, that is, the decreasing of SYM-h lower than -50 nT during its recovery phase toward a quiet condition. This results in higher performance, as it is easier to also forecast such a decreasing during the recovery phase of a solar storm rather than just its onset, which is precisely what LSTM does when not having the value of SYM-H as input. From an operational point of view, it is also worth mentioning that the solar wind data used in the analysis were artificially time shifted as if they had been acquired at the same location as the geomagnetic indices. However, if the tool worked in real time in forecasting the value of SYM-H in the hour following the acquisition of solar measurements at L1, the time required for the solar plasma to reach Earth, i.e., 30 minutes (60 minutes) for a bulk speed of 800 (400) km s^{-1}, would also have to be accounted for. This means that the prediction of geomagnetic disturbances could be provided from 1.5 up to 2 hr before the onset of the geomagnetic storm (depending on solar wind velocity) when the SYM-H is not used as an additional feature in the model.

ORCID iDs

Sabrina Guastavino https://orcid.org/0000-0001-7047-1148
Katsiaryna Bahamazava https://orcid.org/0000-0001-6645-5276
Emma Perracchione https://orcid.org/0000-0003-2663-7803
Gianluca Audone https://orcid.org/0000-0001-9687-4711
Daniele Telloni https://orcid.org/0000-0002-6710-8142
Roberto Susino https://orcid.org/0000-0002-1017-7163
Gianalfredo Nicolini https://orcid.org/0000-0002-9459-3841
Michele Piana https://orcid.org/0000-0003-1700-991X
Anna Maria Massone https://orcid.org/0000-0003-4966-8864

References

Barnes, G., Leka, K. D., Schrijver, C. J., et al. 2016, ApJ, 829, 89
Bloomfield, D. S., Higgins, P. A., McAteer, R. J., & Gallagher, P. T. 2012, ApJL, 747, L41
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Breiman, L. 2001, Mach. Learn., 45, 5
Campi, C., Benvenuto, F., Massone, A. M., et al. 2019, ApJ, 883, 150
Camporeale, E., Wing, S., Johnson, J., Jackman, C. M., & McGranaghan, R. 2018, SpWea, 16, 2
Cander, L. R., & Mihajlovic, S. J. 1998, JGR, 103, 391
Chapman, S., & Ferraro, V. C. A. 1931, TeMAE, 36, 77
Dungey, J. W. 1961, PhRvL, 6, 47
Fairfield, D. H., & Cahill, L. J. J. 1966, JGR, 71, 155
Georgoulis, M. K., Bloomfield, D. S., Piana, M., et al. 2021, JSWSC, 11, 39
Georgoulis, M. K., Yardley, S. L., Guerra, J. A., et al. 2024, AdSpR, in press
Guastavino, S., Marchetti, F., Benvenuto, F., Campi, C., & Piana, M. 2022a, A&A, 662, A105
Guastavino, S., Marchetti, F., Benvenuto, F., Campi, C., & Piana, M. 2023a, FrASS, 9, 399
Guastavino, S., Piana, M., & Benvenuto, F. 2022b, IEEE Trans. Neural Networks Learn. Syst., 35, 1993
Guastavino, S., Piana, M., Tizzi, M., et al. 2022c, NatSR, 12, 20049
Guastavino, S., Candiani, V., Bemporad, A., et al. 2023b, ApJ, 954, 151
Hanssen, A., & Kuipers, W. 1965, On the Relationship between the Frequency of Rain and Various Meteorological Parameters: (With Reference to the Problem of Objective Forecasting) (Koninklijk: Nederlands Meteorologisch Instituut)
Heidke, P. 1926, GeoAn, 8, 301
Hochreiter, S., & Schmidhuber, J. 1997, Neural Comput., 9, 1735
Hu, A., Camporeale, E., & Swiger, B. 2023, SpWea, 21, e2022SW003286
Hu, A., Shneider, C., Tiwari, A., & Camporeale, E. 2022, SpWea, 20, e2022SW003064
Kaneko, H. 2022, QBS, 6, 6
King, J. H., & Papitashvili, N. E. 2005, JGRA, 110, A02104
Kingma, D. P., & Ba, J. 2015, arxiv:1412.6980
Marchetti, F., Guastavino, S., Piana, M., & Campi, C. 2022, PatRe, 132, 108913
Piana, M., Emslie, A. G., Massone, A. M., & Dennis, B. R. 2022, Hard X-Ray Imaging of Solar Flares (Berlin: Springer)
Pricopi, A.-C., Paraschiv, A. R., Besliu-Ionescu, D., & Marginean, A.-N. 2022, ApJ, 934, 176
Pulkkinen, T. 2007, LRSP, 4, 1
Singh, T., Benson, B., Raza, S. A. Z., et al. 2023, ApJ, 948, 78
Telloni, D., Antonucci, E., Bemporad, A., et al. 2019, ApJ, 885, 120
Telloni, D., Carbone, F., Antonucci, E., et al. 2020a, ApJ, 896, 149
Telloni, D., Zhao, L., Zank, G. P., et al. 2020b, ApJL, 905, L12
Telloni, D., Schiavo, M. L., Magli, E., et al. 2023, ApJ, 952, 111
Vourlidas, A., Patsourakos, S., & Savani, N. P. 2019, RSPTA, 377, 20180096
Webb, D. F., & Howard, T. A. 2012, LRSP, 9, 3