Solar Physics (2024) 299:98
https://doi.org/10.1007/s11207-024-02340-9

RESEARCH

Forecasting the Dst Index with Temporal Convolutional Network and Integrated Gradients

Junyan Liu[1] · Chenglong Shen[2,3] · Yang Wang[4,5] · Mengjiao Xu[6] · Yutian Chi[6] · Zhihui Zhong[1] · Dongwei Mao[1] · Zhiyong Zhang[1] · Can Wang[1] · Jiajia Liu[2,3] · Yuming Wang[2,3,7]

Received: 3 January 2024 / Accepted: 20 June 2024 / Published online: 5 July 2024

Abstract

The Disturbance Storm Time (Dst) Index stands as a crucial geomagnetic metric, serving to quantify the intensity of geomagnetic disturbances. The accurate prediction of the Dst index plays a pivotal role in mitigating the detrimental effects caused by severe space-weather events. Therefore, Dst prediction has been a long-standing focal point within the realms of space physics and space-weather forecasting. In this study, a Temporal Convolutional Network (TCN) is deployed in tandem with the Integrated Gradient (IG) algorithm to predict the Dst index and scrutinize its associated physical processes. With these two components, our model can give the contribution of each input parameter to the outcome along with the forecast. The TCN component of our model utilizes interplanetary observational data, encompassing the vector magnetic field, solar-wind velocity, proton temperature, proton density, interplanetary electric field, and other relevant parameters for forecasting Dst indices. Despite the disparity in test sets, our model's forecast accuracy approximates the error levels of the prior models. Remarkably, the prediction error of these machine-learning models has become comparable to the inherent error between the Dst index itself and the actual ring-current strength.

To understand the physical process behind the forecasting model, the IG algorithm was applied in our prediction model, in an attempt to analyze the underlying physical process of the machine-learning black box. In the temporal dimension, it is evident that the more recent the time, the more substantial the influence on the final prediction. Regarding the physical parameters, besides the historical Dst index itself, the flow pressure, the z-component of the magnetic field, and the proton density all significantly contribute to the final prediction. Additionally, IG attributions were analyzed for subsets of data, including different Dst-index ranges, different observation times, and different interplanetary structures. Most of the subsets exhibit an IG matrix with deviations from the mean distribution, which indicates a complex nonlinear system and sensitivity of the prediction to input values. These analyses align with physical reasoning and are in good agreement with previous research. The results affirm that the TCN+IG technique not only enhances space-weather forecast accuracy but also advances our comprehension of the underlying physical processes in space weather.

Keywords Space-weather forecast · Machine learning · Geomagnetic storm

Extended author information available on the last page of the article

1. Introduction

Geomagnetic storms are significant disruptions in the Earth's magnetic field and have exerted profound influences on human life (Gonzalez et al. 1994; Tsurutani et al. 1997; Echer, Gonzalez, and Alves 2006). Several relevant cases have demonstrated that super geomagnetic storms have significant influences on our society, particularly on modern high-tech infrastructure such as spacecraft and radio communication (Heirtzler, Allen, and Wilkinson 2002; Daly 2005; Pulkkinen et al. 2012; Mandea and Chambodut 2020). On 1 March 1989, a severe geomagnetic storm struck power systems in North America, resulting in a 9-h power failure across the majority of the Quebec region (Molinski, Feero, and Damsky 2000). For a 1989 Quebec-like event, the global economic impacts would range from USD 2.4 to 3.4 trillion over a year and lead to a loss in global Gross Domestic Product (GDP) ranging from 3.9 to 5.6% (Schulte In Den Bäumen et al. 2014). The "Halloween" event on 30 October 2003, also constituted a severe disaster impacting the power-system infrastructure, the aviation industry, and satellite communications around the world (Wang et al. 2005; Mannucci et al. 2005; Kappenman 2005; Gopalswamy et al. 2005). On 4 February 2022, 38 Starlink satellites were also destroyed by the geomagnetic storm, and the economic loss is expected to be several tens of millions of dollars (Hapgood, Liu, and Lugaz 2022; Dang et al. 2022). Consequently, the precise prediction of geomagnetic activity has become a necessary subject in the field of space physics.

The Dst index is a pivotal geomagnetic parameter employed to quantify the magnetic turbulence at a temporal resolution of 1 h (Sugria and Chapman 1961). It is defined as the difference between the horizontal component of the measured magnetic field and the horizontal component of the corresponding quiet geomagnetic field (Sugria and Chapman 1961; Sugiura 1963). The global Dst index is derived from the amalgamation of data collected at four low-latitude magnetic observatories: the Hermanus, Kakioka, Honolulu, and San Juan observatories (Nose et al. 2015). The longitudes of these observatories are roughly uniformly distributed in the whole longitude line and these observatories are all located in low latitudes that ensure minimal interference from space electrojets. As a consequence, the Dst index serves as a widely recognized metric for characterizing the strength of the equatorial ring current and quantifying the intensity of magnetic storms.

Over the course of several decades, numerous methods have been proposed to forecast the Dst index. Some researchers (Yermolaev and Yermolaev 2002; Wang et al. 2003; Gonzalez and Echer 2005; Zhao et al. 2011; Shen et al. 2017) used statistical relations between the Dst index and the solar-wind parameters for prediction. Concurrently, others (Burton, McPherron, and Russell 1975; Perreault and Akasofu 1978; Akasofu 1981; Wei et al. 2007; Cai, Ma, and Zhou 2010; Boynton et al. 2011) have utilized inferred coupling functions linking solar-wind dynamics with magnetospheric responses.

In recent years, with the development of computer science, machine learning has made great breakthroughs across various disciplines, including solar physics (Asensio Ramos et al. 2023). In the domain of machine-learning-based Dst-index prediction, Lundstedt and Wintoft (1994) introduced one of the early prediction models, utilizing a neural network. Afterward, Gleisner, Lundstedt, and Wintoft (1996) enhanced prediction accuracy by introducing a time-delay neural network. In a separate approach, Wu and Lundstedt (1997) harnessed a kind of recurrent neural network (RNN) (Elman 1990) to forecast the Dst index, demonstrating the effectiveness of RNN structures in predicting geomagnetic activity. In recent decades, the advancement of computer hardware has catalyzed substantial progress within the field of machine learning. A plethora of machine-learning methodologies have emerged for Dst-index prediction. Various deep-learning models (Lazzús et al. 2017; Lethy

et al. 2018; Xu et al. 2020; Hu, Camporeale, and Swiger 2023) have been proposed, yielding remarkable levels of prediction accuracy.

Despite the remarkable accuracy achieved by deep-learning models in Dst prediction, the inherent opacity of machine-learning models remains a challenge (Buhrmester, Münch, and Arens 2021). The high predictive accuracy of these models often lacks a direct physical explanation. In this study, we have employed a methodology with the combination of the Temporal Convolutional network (TCN) (Bai, Kolter, and Koltun 2018) and the Integrated Gradient (IG) (Sundararajan, Taly, and Yan 2017). The TCN architecture is widely adopted in the field of weather forecasting and has demonstrated substantial enhancements in forecast accuracy (Hewage et al. 2020). The Integrated Gradient (IG) (Sundararajan, Taly, and Yan 2017) is a widely used explainable artificial intelligence (AI) algorithm to look into the machine-learning black box. Utilizing a combination of TCN and IG models, we endeavored to predict the Disturbance Storm Time (Dst) index. Our results reveal that the model demonstrates a noteworthy degree of prediction accuracy. Simultaneously, the model enables a comprehensive analysis of the factors contributing to its accuracy. This paper is divided into five sections. Section 2 is a brief introduction to the calculation algorithm used in this paper. Section 3 introduces the data and the model structure used for training and analysis. Section 4 discusses the results of our model. Section 5 presents the discussion and summary.

2. Computational Method

2.1. TCN Architecture

The Temporal convolutional network (TCN) (Bai, Kolter, and Koltun 2018) draws inspiration from the convolutional architectures designed for sequential data, embodying simplicity, autoregressive prediction, and extended memory capabilities. In comparison to other prevalent generic recurrent architectures, such as Long Short Term Memory (LSTM) (Hochreiter and Schmidhuber 1997) and Gated Recurrent Networks (GRU) (Chung et al. 2014), the TCN architecture demonstrates superior accuracy while maintaining a more straightforward and transparent design. Unlike LSTM and GRU, TCN employs distinct convolutional layers for each step within the interval layers, eschewing the need for a recurrent module that traverses through time. Despite the resultant increase in computational workload, this modification enhances the overall structure's parallelizability, ensuring that the execution time on Graphics Processing Units (GPUs) remains unaffected. Furthermore, this alteration simplifies both the input and output processes, rendering them more straightforward and amenable to analysis.

The core of the TCN algorithm is dilated causal convolution with residual structure. Residual blocks are skip-connection blocks that learn residual functions concerning the layer inputs, instead of learning unreferenced functions. This concept was originally introduced as a fundamental aspect of the ResNet architecture (He et al. 2016). The schematic representation of the residual block's design is depicted in Figure 1. Each residual block contains a branch leading out to a series of transformations $\mathcal{F}$, whose outputs are added to the input $\mathbf{x}$ of the block.

$$o = Activation(\mathbf{x} + \mathcal{F}(\mathbf{x})). \tag{1}$$

This mechanism significantly facilitates the learning of modifications to identity mapping, a phenomenon that has been empirically demonstrated to be advantageous for very deep

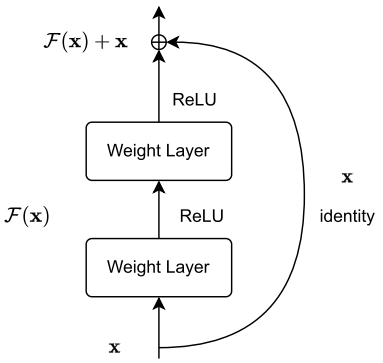

Figure 1 The residual block architecture.

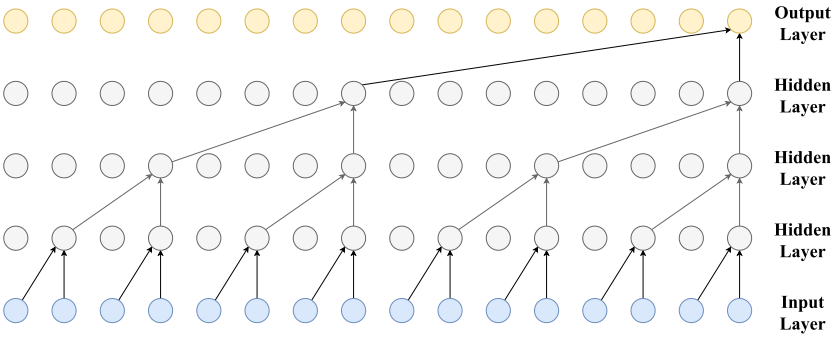

Figure 2 An example TCN structure with dilation factor list [1, 2, 4, 8].

networks. In models incorporating residual structures, accuracy exhibits an increasing trend with greater depth and does not exhibit a deterioration, even under less favorable circumstances. Consequently, the incorporation of residual structures finds widespread application in various machine-learning algorithms, including the TCN structure employed in this study.

Regarding the internal structure within the residual block, the TCN model incorporates a sequence of dilated causal convolutions. Dilated convolution, also known as Atrous Convolution or convolution with holes, is a specialized convolutional technique employed to broaden the receptive field of neural-network models. It achieves a judicious balance between an extensive historical context and reduced computational load. The key idea behind dilated convolution is to expand the convolutional kernel, which in turn skips some of the input points. Causal convolution, on the other hand, constitutes a convolutional structure where the output at time t is exclusively associated with elements from time t and earlier in the preceding layer. This constraint is indispensable for sequence models since, in practical applications, accessing information from the future is unattainable.

With the combination of dilated convolution and casual convolution, the architecture of the TCN algorithm in a specific residual block is shown in Figure 2.

2.2. Integrated Gradient Method

Integrated Gradient (IG), as introduced by Sundararajan, Taly, and Yan (2017), stands as an axiomatic model interpretability algorithm. This method assigns importance scores to

individual input features by approximating the integral of gradients of the model's output concerning the inputs. Renowned for its efficacy, IG serves as a valuable algorithm facilitating human comprehension of the pivotal features influencing the ultimate prediction. In this paper, the IG algorithm is employed within our model to provide an explanatory framework for the rationale behind the predictions.

In the realm of machine learning, a machine-learning model can be conceptualized as a sophisticated function, wherein the input parameters serve as independent variables and the prediction results as dependent variables. Therefore, to scrutinize the contributions of all inputs to the final result, one can utilize the derivative gradient for different inputs as a crucial reference (Baehrens et al. 2009; Simonyan, Vedaldi, and Zisserman 2014). Inputs characterized by higher gradients possess greater significance and play a more foundational role in the prediction process. While this method is straightforward and valuable, it harbors a conspicuous drawback. On occasions, the machine-learning model may exhibit flattening at the input, resulting in zero gradients, despite variations in the function value at the input compared to that at the baseline (Shrikumar, Greenside, and Kundaje 2019). To address this limitation, numerous algorithms have been introduced, with the Integrated Gradients (IG) method emerging as one of the most effective solutions.

The inaccuracy of the classical gradient method comes from the fact that we only calculate the gradient of a single point. Hence, the IG calculation function can be written as the function below:

$$IG_i(x) = (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} \mathrm{d}\alpha. \tag{2}$$

The Integrated Gradients (IG) algorithm assesses the linear trajectory in high dimensions from the baseline x' to the input x and calculates gradients at all points along this trajectory. Its implementation is efficient, requiring only a minimal number of calls to the gradient operator. Furthermore, it is versatile, applicable across various deep networks, and is substantiated by robust theoretical justification (Lundstrom, Huang, and Razaviyayn 2022).

In Equation (2), the attribution calculation is multiplied by the input, which contributes to sharpening attribution maps. Ancona et al. (2017) argue the multiplication with the input has a more fundamental justification, which allows us to distinguish attribution methods in two broad categories: global attribution methods, that describe the marginal effect of a feature on the output with respect to a baseline and; local attribution methods, that describe how the output of the network changes for infinitesimally small perturbations around the original input. Local attributions aim to explain how the input should be changed in order to obtain a desired variation in the output. On the contrary, global attributions should be used to identify the marginal effect that the presence of a feature has on the output. In this study, both global and local attributions are computed and examined to reveal underlying physical relationships.

2.3. Focal Regression Loss Function

The concept of focal loss was initially introduced by Lin et al. (2020), specifically for object detection. The primary aim of Focal Loss is to mitigate the influence of correctly classified instances on the training procedure and redirect attention toward challenging-to-classify examples. Subsequently, Yang et al. (2021) proposed a regression variant of focal loss known as "Focal-R". Given that our Dst-prediction problem is inherently a regression challenge, we employ the "Focal-R" loss function to address the data imbalance in Dst indices.

In the Focal-R loss function, a continuous scaling factor is employed to map the loss within the [0,1] range. More precisely, for the focal loss based on L_1 distance, the ultimate loss is computed as follows:

$$e_{\text{focal}} = \frac{1}{n}\sum_{i=1}^{n}\sigma(|\beta e_i|)^{\gamma} e_i, \tag{3}$$

where e_i is the L_1 error for the ith sample, $\sigma(\cdot)$ is the sigmoid function, and β, γ are hyperparameters. Through the application of the focal-R loss, the impact of accurately predicted data on the training process is minimized, thereby shifting the focus of the training towards challenging data.

2.4. Sequential Model-Based Algorithm Configuration

The Sequential Model-based Algorithm Configuration (SMAC) was initially introduced by Hutter, Hoos, and Leyton-Brown (2011) and subsequently enhanced by Lindauer et al. (2022). It is one of the automated machine-learning methods for hyperparameter optimization and has been widely used in many fields.

Within the realm of machine learning, hyperparameter optimization addresses the challenge of identifying an optimal set of hyperparameters for a given learning algorithm. Among various algorithms, Bayesian optimization emerges as a widely adopted approach. This method intelligently navigates the hyperparameter space, determining the next combination to explore based on prior observations. In contrast to grid search and random search, Bayesian optimization excels in yielding superior results with fewer evaluations, owing to its capacity to assess experiment quality before execution. While the conventional Bayesian optimization approach focuses exclusively on tuning numerical parameters, the SMAC algorithm is uniquely tailored to handle categorical parameters.

The fundamental concept behind Bayesian optimization involves constructing a model that estimates the metric value for a given set of hyperparameters. In the case of the SMAC algorithm, this is achieved through a random forest model capable of simultaneously handling both categorical and numerical parameters. Leveraging the results derived from the preceding hyperparameters, the random forest formulates a novel set of hyperparameters associated with the highest estimated score. Subsequently, the model undergoes retraining using this updated outcome, initiating the iterative process once more. This iterative approach expedites the discovery of superior hyperparameters earlier and faster.

2.5. Stacking Algorithm

The stacking algorithm, introduced by Wolpert (1992), stands out as one of the widely employed ensemble learning techniques aimed at enhancing model performance. The ensemble technique is the idea that combines the predictions of several base estimators built with a given learning algorithm to improve the single estimator. This technique has found extensive applications across various domains of machine learning, including the field of space-weather forecasting (Xu et al. 2020; Yang et al. 2023). Returning to the stacking algorithm, this method involves the amalgamation of multiple individual models, effectively mitigating prediction bias. More precisely, predictions from each individual estimator undergo a stacking process, serving as input for a final estimator that computes the ultimate prediction. This approach allows the training of multiple models addressing the same problem, and leveraging their combined output, a new model is constructed, showcasing improved

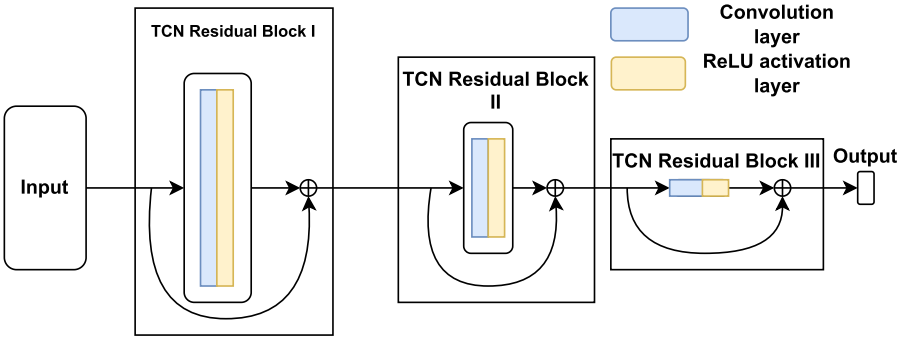

Figure 3 The TCN network structure used for Dst-index prediction in this paper.

performance. Consequently, the stacking algorithm is implemented on the best-performing models, culminating in the creation of an enhanced prediction model. The resultant stacked model boasts higher accuracy, serving as the foundation for the conclusive analysis.

3. Dataset and Prediction Model Details

The Dst-index data and solar-wind data used in this paper are obtained from NASA's OMNI database (King and Papitashvili 2005). The data that we used contain the hourly Dst-index and solar-wind data from 00 UT on 1 January 1963 to 23 UT on 31 December 2016 since the final index with the final quality check is now available only up to 2016. The interplanetary parameters considered in this paper include the interplanetary magnetic field (B_x, B_y, and B_z), the solar-wind velocity (v_x, v_y, and v_z), electric field (E), proton density (n_p), temperature (T), and dynamic pressure (P_dyn). The interplanetary data presented herein are collected at the Lagrange L1 point in space through satellite observations. Unfortunately, a significant portion of interplanetary data from the early years is absent due to a lack of observations. To address this journal, we simply discard this part of the data. Consequently, we have compiled a dataset consisting of 230,527 sets of input and output data pairs, which are utilized for prediction and analysis purposes.

In our study, we categorize the dataset based on monthly observations, taking into account variations in solar-activity intensities across different years. The training set comprises observation data from January, March, May, July, September, and November, serving as input for updating model parameters through training functions. Conversely, the validation set consists of data from April, August, and December and is employed to select the optimal model, mitigating the risk of overfitting. Additionally, the testing set is constructed from data obtained in February, June, and October. This set is utilized to evaluate the Dst-prediction model and conduct theoretical analyses. Notably, each year's data is subdivided into three parts, ensuring that all three subsets encapsulate both the solar-maximum and -minimum periods of distinct solar cycles. This strategic division ensures a uniform distribution across subsets, enhancing the model's performance in addressing real-world space-weather forecasting challenges.

Regarding the model architecture, we adopt the TCN model concept, incorporating multiple dilated causal convolution layers as residual blocks. The schematic representation of our model is presented in Figure 3. Each residual block in our model consists of a convolution layer followed by a ReLU activation layer. Notably, the final prediction model is

composed of three interconnected residual blocks with varying widths. In addition to these TCN residual blocks, a linear layer is appended to the end of our model to ensure that the model's output conforms to the desired prediction shape ($N \times 1$).

The lengths of input and output sequences are determined through a series of model selections. Employing the Neural Network Intelligence (NNI) software (Microsoft 2021), the SMAC algorithm generates a set of model hyperparameters, which are then utilized to predict outcomes in the validation dataset. The NNI software is an automated machine-learning software and is mainly used for hyperparameter optimization with the SMAC algorithm in our Dst-index prediction research. The optimal model, yielding the highest prediction accuracy, employs a 57-h input and 1-h output. This optimal set of hyperparameters is utilized for subsequent statistical analyses. Consequently, the prediction model in this study utilizes interplanetary data from the present (t) to 56 h prior ($t = 56$ h) to forecast the Dst index 1 h ahead ($t + 1$ h). These parameters are derived entirely from data analysis but are also underpinned by a physical basis, as elaborated in Section 4.2.

In the model-training part, diverse techniques were employed to enhance the accuracy of our model. Since most of the time, there is no geomagnetic storm, achieving precision in training a magnetic-storm prediction model poses a formidable challenge. This challenge aligns with the widely recognized issue in machine learning referred to as data imbalance (He and Garcia 2009; Branco, Torgo, and Ribeiro 2015; Krawczyk 2016). The trained model often tends toward predicting the absence of a magnetic storm regardless of the input. To mitigate the impact of imbalanced data, the focal-R loss function is used during our training process. By implementing the focal-R loss, emphasis is deliberately reduced on data points with higher accuracy, redirecting attention towards less accurate data. This strategic adjustment significantly mitigates the impact of data imbalance.

In addition to using the focal-R loss function, the stacking strategy is additionally employed to enhance prediction accuracy. Since the SMAC algorithm generates several different hyperparameters, it is highly suitable for using ensemble learning to enhance prediction accuracy. Stacking has played a pivotal role in further elevating the precision of our Dst-index prediction model. In the process of constructing and training our model, we employed specific Python libraries to streamline our code and conduct data analysis. The utilized libraries are enumerated below: numpy (Harris et al. 2020), pandas (The pandas development team 2020), matplotlib (Hunter 2007), scipy (Virtanen et al. 2020), sklearn (Pedregosa et al. 2011), Pytorch (Paszke et al. 2019), etc.

4. Results

4.1. Prediction Accuracy

Considering the complexity of the prediction problem, it is difficult to describe how good the pros and cons of a model are. Therefore, we use three different evaluation functions to analyze our result. These evaluation functions are the mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2).

$$\mathrm{MAE}(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} |y_i - \hat{y}_i| \quad , \tag{4}$$

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=0}^{N} (y_i - \hat{y}_i)^2} \quad , \tag{5}$$

$$\mathrm{R}^2(y,\hat{y}) = \frac{Cov^2(y,\hat{y})}{Var(y)Var(\hat{y})} = \frac{\left(\sum_{i=1}^{N}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\right)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2\sum_{i=1}^{N}(\hat{y}_i-\bar{\hat{y}})^2} \quad . \tag{6}$$

Here, y and $\hat{y}$ represent the predicted and actual Dst indices for the entire dataset, y_i and $\hat{y}_i$ pertain to the predicted and actual Dst indices for the ith data point, $\bar{y}$ and $\bar{\hat{y}}$ denote the average values for the predicted and actual Dst indices and N signifies the number of data points in the dataset. Each of these three evaluation functions contributes uniquely to quantifying the accuracy of a prediction model and all have their own characteristics. The MAE and RMSE evaluation functions share the same units as the input parameters, whereas the correlation coefficient is a scale-free score. In the context of a highly accurate model, the calculated MAE and RMSE between predicted and actual values tend to be smaller, while the correlation coefficient tends to be larger. In an ideal scenario where the predicted value precisely aligns with the real value, both the MAE and RMSE register as zero, and the correlation coefficient (R) attains a value of 1. For all our test set data, the MAE value is 2.12 nT, the RMSE value is 3.09 nT, and the R-squared is 0.973.

Through these three different parameters, the performance of the prediction model can be quantified, but it is still very difficult to compare our results with other previous models because the test dataset is not exactly the same. Furthermore, due to the limitations of these evaluation methods, their sensitivity to different data is different. In general, the data to be predicted with greater absolute value will have larger RMSE and MAE, and higher correlation coefficient. For our Dst-prediction problem, due to the relatively quiet solar activity during the Solar Cycle 24, the previous testing set using only part of the Solar Cycle 24 observation data usually has a smaller RSME and MAE value.

Although it is difficult to compare among different people's works, it can be verified from another perspective that we all have obtained a very high accuracy. Although the Dst index is the most essential and widely used parameter to quantify the Earth's global magnetic disturbance, it is still slightly different from the actual ring-current strength. The Dst index usually underestimates the solar activity and its bias ranges from -15 nT at solar maximum to zero at solar minimum (Lühr and Maus 2010; Temerin and Li 2015). In addition, the semiannual variation in the Dst index is excessively large compared to all other indices of geomagnetic activity (Cliver, Kamide, and Ling 2000). It is unrelated to the development of storms or the ring current and its approximate value is about 2.5 nT (Mursula and Karinen 2005). Therefore, it is meaningless to pursue high accuracy in Dst-index prediction, since it is slightly different from the actual physical process. Moreover, all those recent prediction models that are based on machine learning have reached the error level between the Dst index and the actual physical process. Table 1 presents the comparison between our model and several different previous models. Considering these models use different testing sets for result analysis, the evolution results can only be used in rough comparison. Notably, all models exhibit exceptionally high accuracy, with prediction errors comparable to the error level between the Dst index and the real physical processes. Consequently, further enhancements in forecast accuracy may be limited. Instead, the research focus should shift towards analyzing how these deep-learning models attain such precise predictions, which forms a crucial aspect of this paper.

In addition to the average bias calculated with the three evaluation functions, we also obtain the relationship between our final Dst-index prediction and its ground-truth value for all the testing sets. The comparison is shown in Figure 4(a). From the scatter plot, we

Table 1 Comparison between our prediction model and several other Dst-forecast models.

Prediction model	The corresponding testing set	RMSE (nT)	R^2
ANN+PSO (Lazzús et al. 2017)	2010.1.1 00:00 – 2016.8.31 23:00	3.57	0.956
LSTM+GP (Gruet et al. 2018)	20% of all dataset (2001.1.14 00:00 – 2016.12.31 23:00)	5.25	0.933
Bagging (Xu et al. 2020)	2016.1.1 00:30 – 2017.5.31 23:30	2.85	0.966
Dst Transformer (Abduallah et al. 2022)	2021.10.1 – 2021.11.15	5.40	0.750
LSTM (Li et al. 2022)	2018.10.21 14:00 – 2020.12.31 23:00	3.02	0.963
Multi-fidelity Boosting (Hu, Camporeale, and Swiger 2023)	20% of 66 selected storm periods with K-fold cross validation	2.59	–
EMD-LSTM (Zhang et al. 2023)	data from 2019 (quiet period)	3.29	0.903
	2023.1.1 – 2023.5.21 (active period)	8.87	0.865
our TCN-IG model	February, June, and October from 1963 to 2016	3.10	0.973

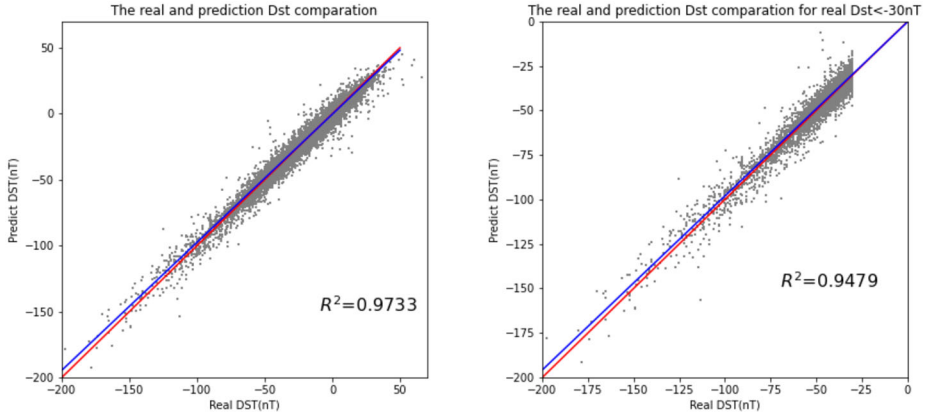

Figure 4 Comparison between the prediction result and real Dst-index values. The red line presents the linear relation result and the blue line is the fitted result. Panel (a) uses all the data in the testing dataset, while Panel (b) only uses those data whose ground truth Dst index is lower than -30 nT.

can see that the predicted and real Dst indices are closely fitted around the $y = x$ function and have a very high correlation coefficient. Considering that the dataset contains quiet geomagnetic-field observations, we then filter the data with Dst-index values lower than -30 nT. The R-squared evaluation function is 0.949, and the scatter plot comparison is shown in Figure 4(b). Our result is still good enough for those periods with a geomagnetic storm. Thus, the prediction result of our TCN model is rather reliable and can be used in space-weather forecasting.

In addition to the overall analysis, the accuracy of different observation data is also analyzed in detail. To verify whether there is overfitting in our model, the three validation metrics are also calculated for each month since our dataset is divided based on months. Figure 5 provides a comparison across different months. The results are similar among all months, indicating that the training is not overfitting and our model is not biased to any specific group of months.

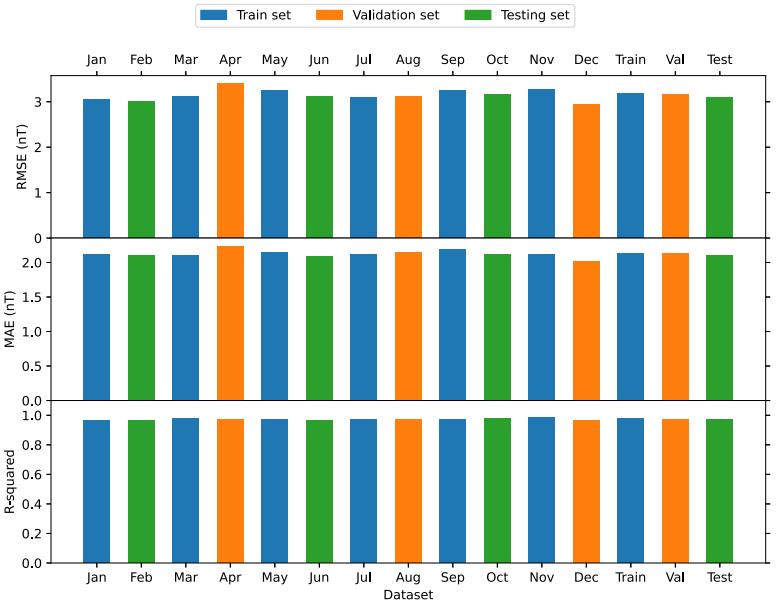

Figure 5 The RMSE, MAE, and R-squared evaluation results for different months and different datasets. It is obvious that different parts of the dataset have a similar accuracy and our model is not biased to any specific group of months.

Correspondingly, we also conducted similar analyses for different years. Figure 6 presents the evaluation-function results of our model's prediction and the ground truth for different years. Due to the long span of the dataset and the fact that some observations are missing in the early years, we only present results for the 17 years of the twenty-first century. As is evident from the figure, the prediction accuracy varies from year to year. This discrepancy arises from the disparate distribution of data within each year and the varying sensitivity of the three evaluation functions to distinct datasets. Therefore, it is more appropriate to divide the dataset using different months rather than years. Moreover, this result implies that the inherent limitations of these three evaluation methods hinder direct comparisons of their values with those of other models, primarily due to variations in the datasets employed across different research papers.

Considering our dataset contains observation data from 5 different solar cycles, the three evaluation functions for different solar cycles can be shown in Table 2. It is easy to see that the accuracy varies across different solar cycles, and dividing the dataset according to different solar cycles is also not good enough. Further analysis using variance calculation reinforces this result. For data from different solar cycles, the variance of the RMSE assessment metric is 0.380 nT. In comparison, the RMSE variance for monthly split data is 0.127 nT, and for yearly split data, it is 0.675 nT. These findings indicate that splitting the dataset based on months is a more appropriate method.

The general result comparison does not seem intuitive, so 6 specific geomagnetic-storm events are chosen as illustrative prediction examples. The comparison between our prediction results and the corresponding ground-truth values is shown in Figure 7. The top four events here are the largest magnetic-storm events selected from the test set without missing observation data. Despite the delayed update of the final Dst index, the analysis was

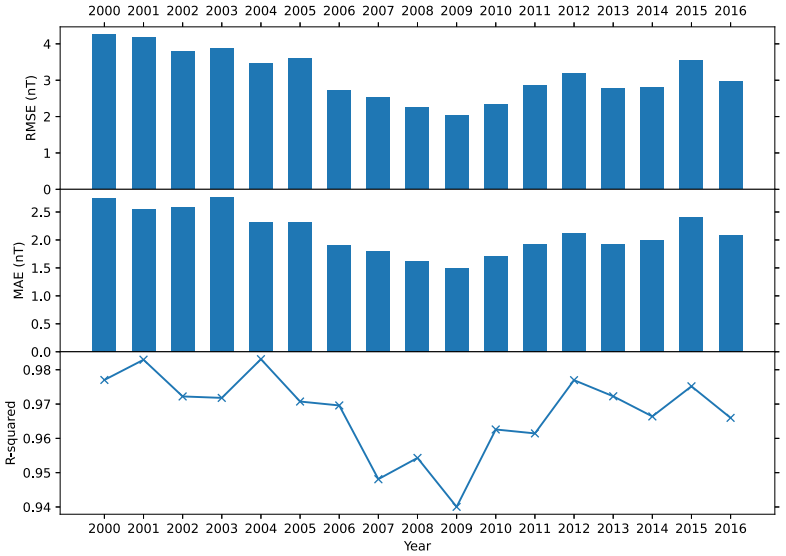

Figure 6 The evaluation-function results of our model's prediction and the ground-truth value during different years. The prediction accuracy varies from year to year due to the disparate distribution of data.

Table 2 The prediction accuracy for different solar cycles.

Solar Cycle	MAE	RMSE	R^2
20	2.29	3.42	0.970
21	2.63	3.72	0.967
22	2.10	2.92	0.973
23	2.16	3.15	0.974
24	1.94	2.88	0.970
all	2.11	3.10	0.973

restricted to data prior to 2016. It has been a long time since 2016, and the space-weather environment is changing all the time. Therefore, two recent storm events from two different solar cycles were added to assess the real-time forecasting capabilities of the prediction model. For these events, the prediction outcomes were evaluated against the real-time or provisional Dst index. Although these data remain subject to ongoing updates and may not be entirely reliable for scientific research, it can still roughly provide an approximation of our model's prediction accuracy. In these figures, the orange lines represent our predicted outcomes during the respective periods, while the red lines depict the actual Dst-index values. Notably, our prediction results exhibit a close alignment with the actual curve at each stage of the magnetic storm. Consequently, our Dst-index prediction model demonstrates a high level of accuracy, rendering it applicable for real-world space-weather forecasting.

4.2. Model Explanation

Given the inherently opaque nature of machine learning, a notable gap exists in the interpretability of the neural networks in prior models. In contrast to existing approaches, some

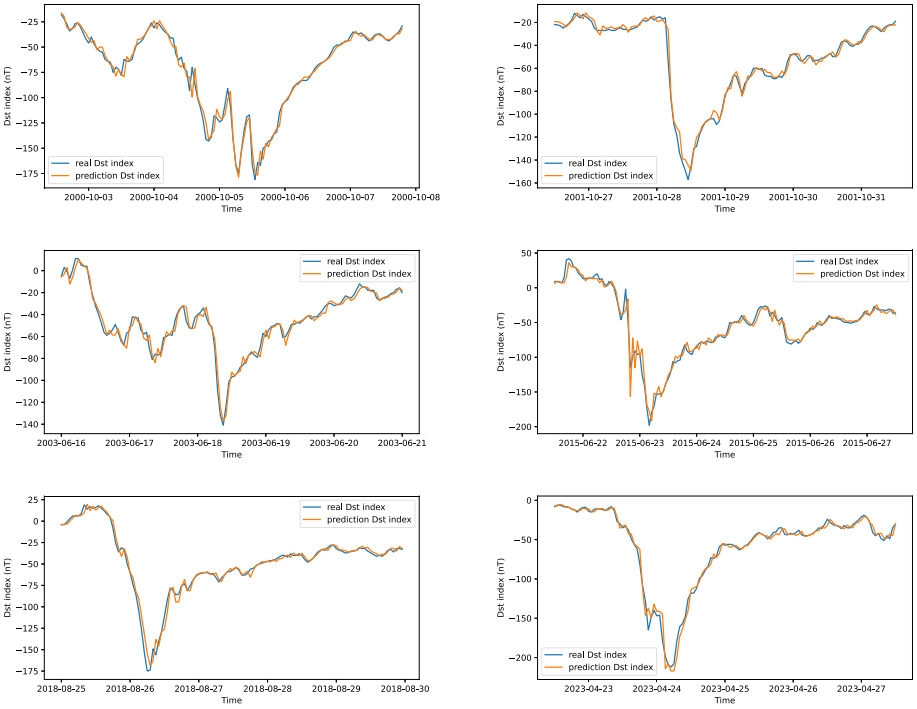

Figure 7 The comparison between the predicted results generated by the TCN network and the actual Dst-index values. The orange lines represent our prediction results, while the red lines depict the real Dst index. Notably, the top four geomagnetic storms are all selected from the testing set and the bottom two events are two recent storms whose Dst-index values have not been fully manually corrected. These magnetic storms all remain unutilized in the training phase, so there are no data-contamination issues. It is noteworthy that, for all these geomagnetic storms, our prediction results closely align with the ground-truth values.

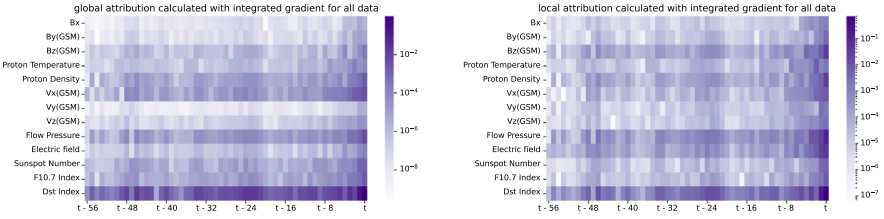

Figure 8 The attribution of different input data calculated with the IG method. The left figure displays the global attribution computed by the IG algorithm, while the right figure shows the local attribution. A larger value (represented by a darker color in the figure) indicates that the corresponding parameter has a more substantial impact on the final Dst-index prediction outcome.

physical analyses are also given along with our prediction model. We can provide an analysis of the contribution of individual input parameters to the forecast results.

To assess the interpretability of our machine-learning model, we utilize the integrated gradient (IG) method. The global and local attribution matrices generated by this method illustrate the influence of various input physical variables on the predicted Dst index, as depicted in Figure 8. To show the results more clearly, the IG attribution matrices are presented

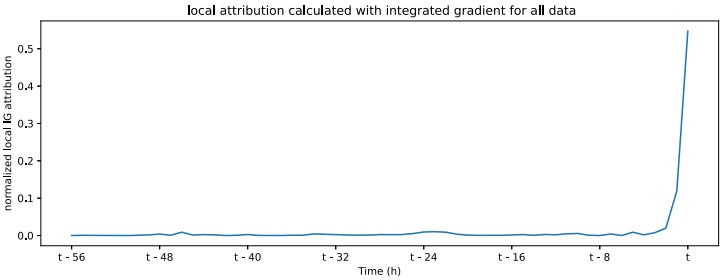

Figure 9 The local attribution for different time interval data calculated with the IG method.

on an absolute logarithmic axis. A lot of common information can be seen from these two figures although they have different emphases. It is evident that the Dst index in the past few hours has the greatest contribution. The solar-wind flow pressure and z-component of the geomagnetic field are all necessary input parameters. These physical analyses will be explored in detail later.

The two-dimensional result is sometimes difficult to analyze horizontally, so we also plotted the contributions of different time intervals and different interplanetary physical parameters to the results separately. Figure 9 shows the local attribution of input data for different time intervals to the final forecast results. From the temporal dimension, most of the attribution comes from the current observation time and those data earlier than 3 h have a very low contribution to the result. This result can also be explained in physics. The interplanetary data can be used for space-weather forecasts mainly because those data are observed at the Lagrange point. It takes several hours to several days for a CME eruption to propagate to Earth, so the observations midway can provide a series of useful forecasts. The common solar-wind velocity ranges from 200 to 750 km s^{-1} and the average speed is approximately 450 km s^{-1}. The distance between the Earth and the Lagrange point L1 is approximately 1.5×10^7 km, so it usually takes one hour for a CME event to propagate from the L1 point to Earth. Thus, our model's outcome that most of the attributions come from the current-hour observation is reasonable. Apart from the input sequence length, the output sequence length selection is also explainable with it. The propagation time between the L1 point and Earth is 1 h on average, so we only predict the Dst index for the next hour, and the result is accurate most of the time.

As for the physical parameter dimension, the local attributions for various input parameters are also calculated. The original value and the absolute value of the attribution are both plotted in Figure 10. The two figures are used to display and analyze different information. Positive values represent a positive contribution to the forecast and negative values represent a negative contribution to the forecast. A larger absolute value represents a greater impact on the final prediction. It is obvious that the past Dst index is the most important parameter and has the greatest contribution to the final prediction. As for other parameters, the solar-wind flow pressure, the z-component of the magnetic field, and the solar-wind proton density all contribute significantly to the forecast. This result can also be explained physically and agrees well with previous research (Russell, McPherron, and Burton 1974; Burton, McPherron, and Russell 1975).

Figure 8 only displays the attribution data for the entire testing set, which includes a diverse range of data. The plot for average value may result in the loss of crucial information, so we also compute the IG matrix for different data subsets. We only calculate the local IG attribution since these data subsets contain different data, and comparisons using global

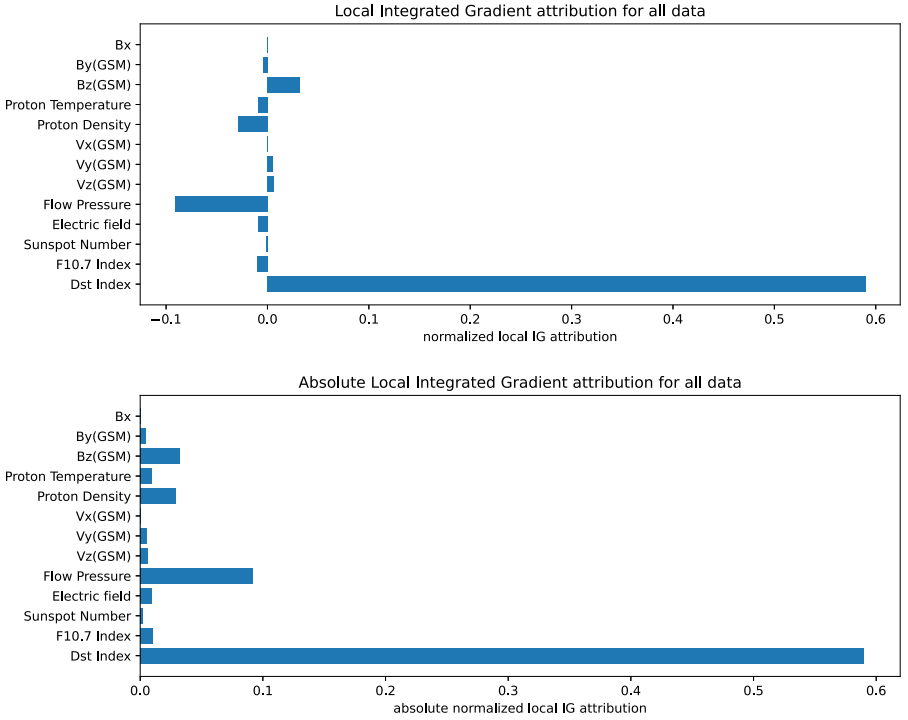

Figure 10 The local attribution for different input parameters calculated with the IG method. The top figure represents the original IG attribution. Positive values represent a positive contribution to the forecast and negative values represent a negative contribution to the forecast. The bottom figure displays the absolute value of the local attribution. A larger absolute value represents a greater impact on the final prediction.

imputation analysis will introduce systematic bias. Figure 11 is the local IG attribution matrix for those data with a Dst index lower than −30 nT and −200 nT. They correspond to the local attribution of data during quiet periods, magnetic-storm periods, and extreme magnetic-storm periods. Obtaining distinct results solely from the calculated matrix remains challenging; hence, the relative local integrated gradient attribution matrices are also computed. The attribution results divided by the attribution results from the whole dataset are shown in Figure 11 with the logarithmic axis. A redder color signifies a higher contribution relative to the base attribution matrix, whereas a bluer color indicates a diminished contribution.

The 2D attribution is still not clear enough, so the attribution for temporal dimension and parameter dimension are also calculated. Figure 12 presents the relative attribution for these three data subsets. Notably, the attribution for those without the geomagnetic-storm period (Dst > -30 nT) aligns closely with the overall attribution for the whole dataset. However, as the Dst index decreases, the attribution for the past 3 – 9 h diverges from the average attribution value. This indicates that during geomagnetic storms, input values from these hours have a distinct contribution compared to quiet periods. In addition, the attributions for input data from even longer ago deviate slightly from the average attribution matrix, which means it makes sense to use earlier data for forecasting. Theoretically, this deviation can be explained by the low correlation between the quiet period and the preceding and succeeding hours. Hence, data from earlier time frames contributes minimally to the forecast. In

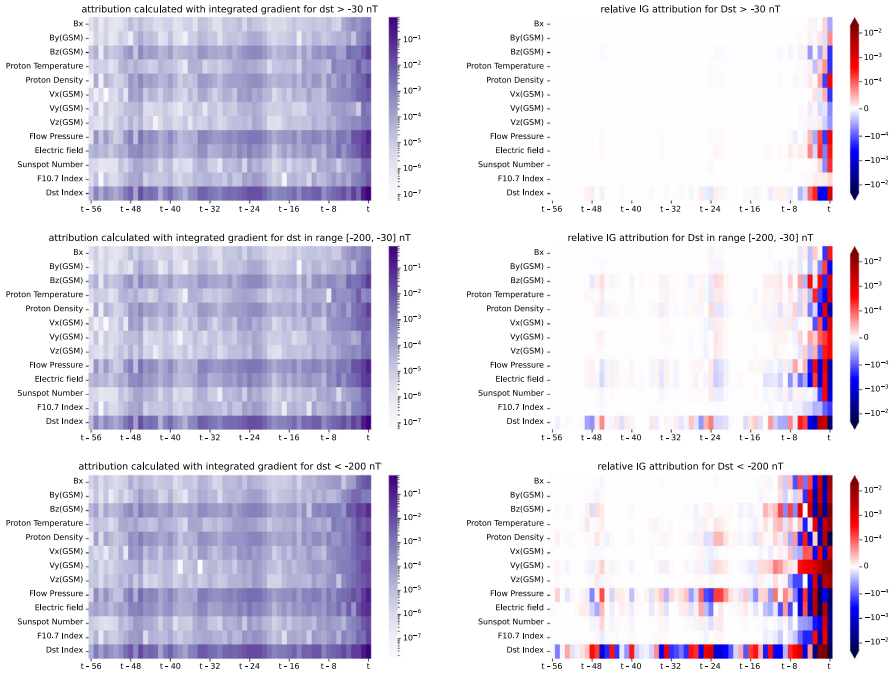

Figure 11 The local integrated gradient attribution matrix for data with Dst index greater than −30 nT, between −30 and −200 nT, and less than −200 nT. The left three figures are the absolute integrated gradient attribution matrices for these three Dst-index ranges. In these figures, the darker color in the figure represents a higher attribution and a greater impact on the prediction result. The right three figures show the subtracted differences of the attribution results from the whole dataset and represent the relative attribution for different input parameters. A redder color indicates a greater contribution relative to the base attribution matrix while a bluer color represents a lower contribution.

contrast, during geomagnetic storms, observations from a few hours prior exhibit a higher correlation, leading to their attributions departing from the whole dataset's average value.

Figure 13 shows the attribution of different interplanetary parameters to the final prediction results for different Dst-index ranges. From these figures, it is evident that the historical Dst index stands out as the most significant parameter. Among those interplanetary parameters, the flow pressure is the second most essential parameter. The z-component of the geomagnetic field and the proton density also have a strong contribution to the forecast. In addition, as the Dst index decreases, the interplanetary magnetic field in the z-direction and the electric field in the y-direction have larger contributions relative to the quiet period, which means that these parameters have a stronger correlation with the geomagnetic storms. This finding aligns with prior research and holds physical explanations. The convective electric field in the magnetosphere is the predominant factor influencing the intensity of the ring current. The solar-wind plasma speed and the z-component of the magnetic field stand out as the two key parameters that affect the electric field. Additionally, in the widely held theoretical explanation of geomagnetic storms, the southward magnetic field in the interplanetary coronal mass ejections or stream-interaction regions reconnects with the Earth's planetary magnetic field at the day-side magnetopause and injects the magnetic and particle energy into the Earth magnetosphere. Thus, the z-component of the interplanetary magnetic field is

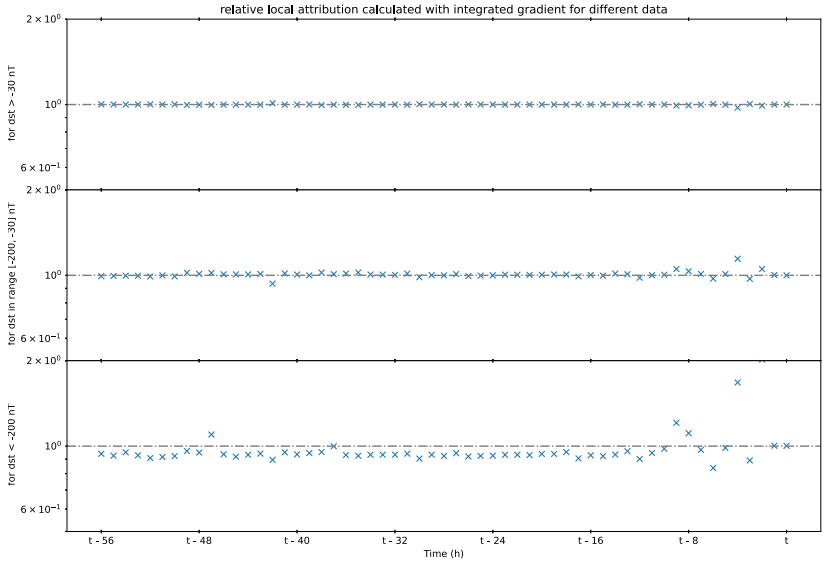

Figure 12 The relative local attribution of different time intervals for data with different Dst-index ranges. The top panel is for data with their Dst index greater than −30 nT. The middle panel is for data between −30 and −200 nT. The bottom panel is for data less than −200 nT. The blue cross represents the relative attribution ratio between the subsets and the whole dataset. The horizontal dotted line represents a relative attribution value of 1, which means that the subset of data is not biased relative to the whole dataset.

the key parameter of geomagnetic-storm forecasting. Our model's result highly agrees with the results from the previous theoretical research.

Since there are two main types of causes of geomagnetic storms: the interplanetary coronal mass ejection (ICME) and the stream interaction region (SIR), the analysis for the prediction for different sources is also conducted. Figure 14 shows the local IG attribution matrix for the final Dst prediction during the ICME and SIR period. The time range of ICMEs is obtained from Chi et al. (2016) and the time range of SIRs is obtained from Chi et al. (2018). In addition, the 2D attribution matrix, the summation for each input parameter during the ICME and SIR period is also calculated and plotted in Figure 15. Based on the comparison presented in Figure 14 and Figure 15, additional implications can be inferred. Parameters, such as the Dst index, the flow pressure, the z-component of the magnetic field, the y-component of the solar-wind velocity, etc., which significantly affect forecasting, exhibit different contributions during the ICME and SIR periods. This indicates that the two sources are physically different and have completely different characteristics.

Given the extended duration of our testing dataset, the analyses for our prediction during different solar cycles have also been conducted. Figure 16 presents the local IG attribution for different input parameters during five different solar cycles. A notable finding from this figure is that most input parameters exhibit similar contributions to the prediction across different solar cycles. Considering the comparison is slightly complex, the standard deviation and the relative standard deviation, which is defined as the ratio of the standard deviation σ to the mean μ (σ/μ), are then calculated and plotted in Figure 17. Analysis of these figures reveals that the x-components of the solar-wind velocity and sunspot number are the primary parameters with varying contributions to the prediction across different solar cycles. The fluctuations in other parameters are minimal relative to their actual values.

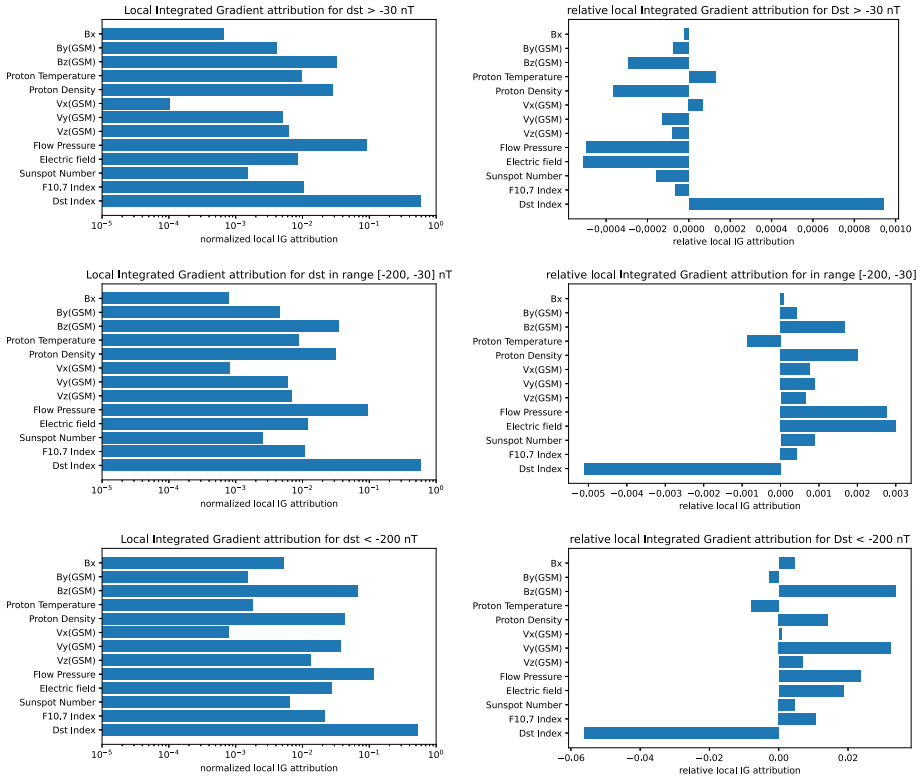

Figure 13 These figures show the local attribution for different interplanetary observation data calculated with the IG method. The three figures on the left are for the data with their Dst index greater than −30 nT, between −30 and −200 nT, and less than −200 nT. The three figures on the right are the relative local IG attribution for different input parameters and different Dst-index ranges. In order to compare the interrelationships more clearly and accurately, all attributions in the figures were calculated and plotted in absolute values. Besides, the three figures on the left are plotted on logarithmic axes to make the comparison clearer.

In addition, the analysis across different solar cycles, the same analytical process was also conducted for different individual years. Figure 18 shows similar plots of the standard variation and the relative standard variation for various years. Figure 18 and Figure 17 exhibit striking similarities. The two most frequently altered physical parameters in the IG attribution are also the x-components of the solar-wind velocity, as well as the sunspot number. The similarity of this result can be attributed to the nature of the forecasting model. These two parameters are more sensitive to the input values and consequently contribute differently in different subdatasets.

In contrast to the similarities observed in Figure 17, Figure 18 exhibits several notable differences. One of the main differences is that the relative standard deviation of the sunspot number becomes lower than that in Figure 17 and the relative variation of the x-component of the solar-wind velocity surpasses that of other parameters. These disparities are attributed primarily to the variability of data across different solar cycles. Given the limited number of solar cycles, statistical analysis may introduce bias. To corroborate this hypothesis, similar statistical analyses were performed on monthly data and are presented in Figure 19. In comparison to Figure 18, Figure 19 yields nearly identical results. This suggests that the

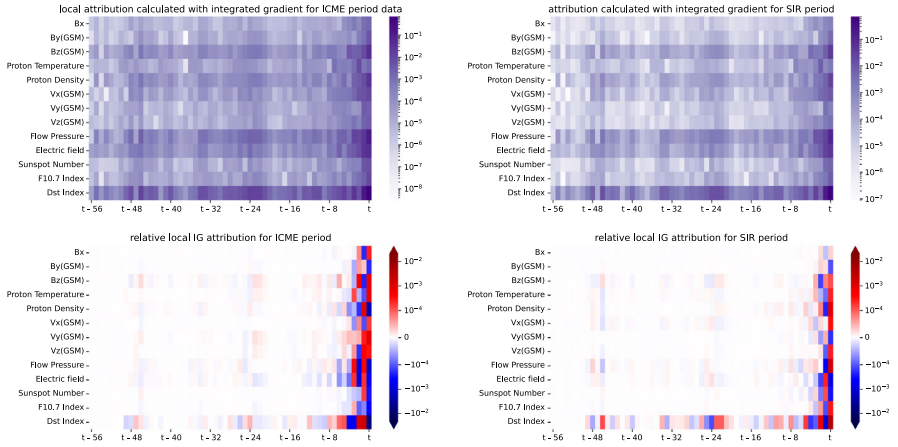

Figure 14 The local IG attribution matrix for those data during the ICME (left) and SIR (right) period. The top two figures show the absolute value matrix and the bottom two figures display the relative value matrix.

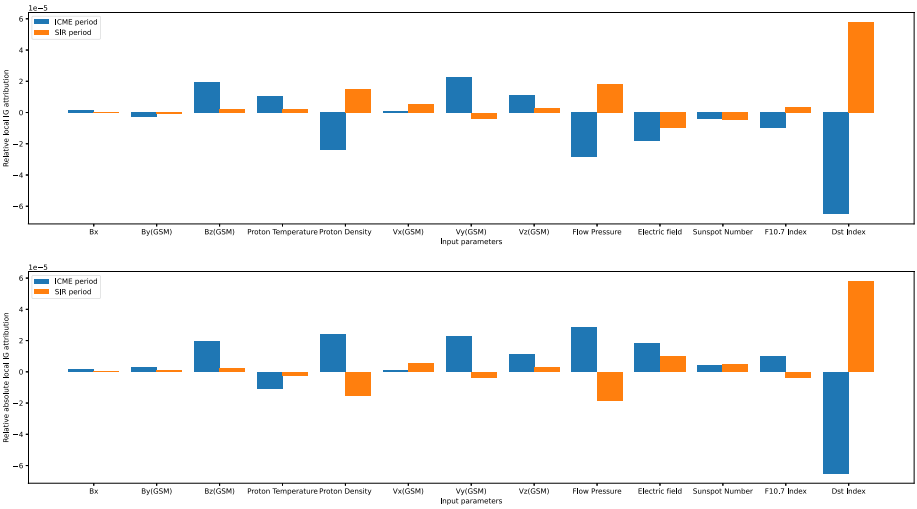

Figure 15 The relative IG attribution matrix for various input parameters during the ICME and SIR period. The relative attribution in the top figure is calculated by subtracting the average attribution of the entire dataset from the attribution of the ICME or SIR event phase. The relative absolute attribution in the bottom figure is similar, except that the calculated parameters are the absolute attribution rather than the origin value.

observed patterns stem from intrinsic characteristics rather than temporal variations in the observations. In other words, the influence of the input parameters on the results remains largely unaffected by the solar cycle, including the solar-maximum and -minimum phases.

To further analyze the attribution of each input physical parameter, the average attribution and average input values for different months are plotted in Figure 20. The left figure depicts the relationship between the local attribution and observation time for each input parameter. The right figure illustrates the average input values for various input parameters over time. Within these two figures, graphs at the same vertical heights represent the same

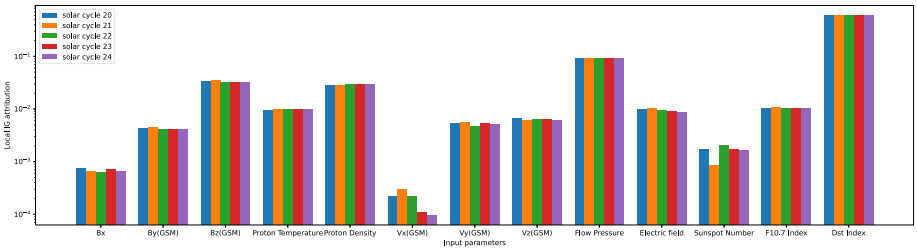

Figure 16 The local Integrated Gradient attribution for different input parameters during five different solar cycles.

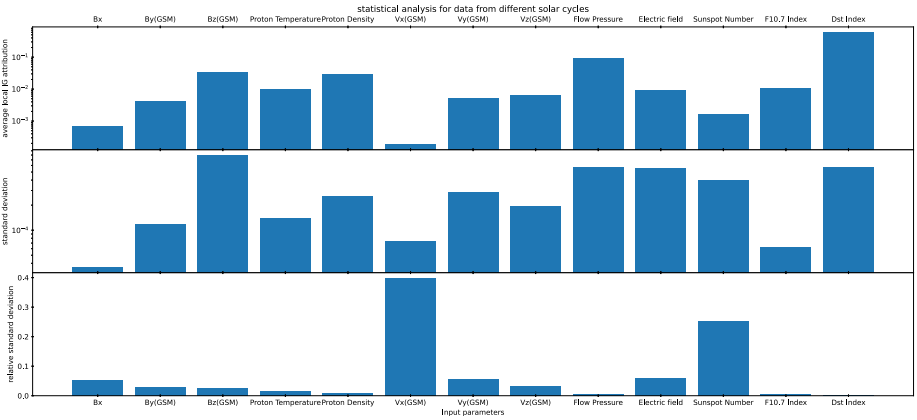

Figure 17 The top panel is the average local IG attribution for the different input parameters. The middle panel is the calculated standard deviation during different solar cycles for the different input parameters. The bottom panel is the relative standard deviation or the coefficient of variation (defined as the ratio of the standard deviation σ to the mean μ, σ/μ) for the different input parameters.

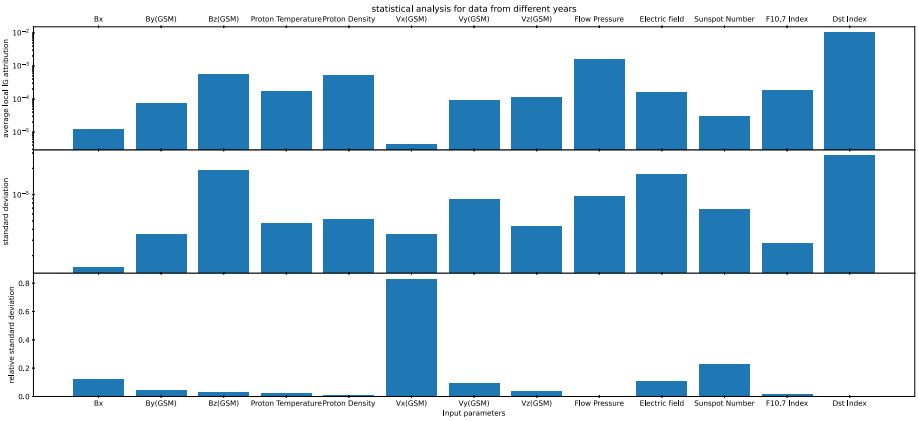

Figure 18 Similar to Figure 17, except that the division of the data subsets is based on years rather than solar cycles. The top panel is the average local IG attribution for the different input parameters. The middle panel is the calculated standard deviation during different years for the different input parameters. The bottom panel is the relative standard deviation or the coefficient of variation.

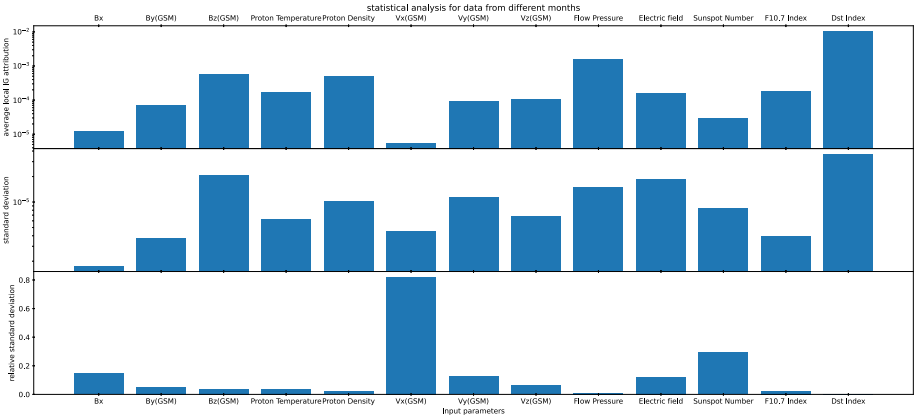

Figure 19 Similar to Figure 17 and Figure 18, except that the division of the data subsets is based on months. The top panel is the average local IG attribution for the different input parameters. The middle panel is the calculated standard deviation during different years for the different input parameters. The bottom panel is the relative standard deviation or the coefficient of variation.

physical input parameter. The vertical solid lines denote the solar minimum, while the vertical dashed lines represent the solar maximum. Comparative analysis indicates that most parameter attributions exhibit a strong relationship with solar-activity intensity, though the relative variation is often very small.

5. Conclusion and Discussion

In the course of this study, we utilized a combination of the TCN model and the IG algorithms to predict the Dst index in the immediate future and scrutinize its implementation process. It becomes more feasible to achieve precise space-weather forecasts by employing the machine-learning method. Notably, the MAE of our model between the prediction 1 h in advance and the ground-truth value is merely 2.12 nT. This precision has already reached the magnitude of error between the Dst index itself and the actual ring-current strength. Nevertheless, the potential for further accuracy improvements is limited. Accordingly, future research should focus on exploring the exceptional predictive capabilities of machine-learning algorithms in the forecasting of the Dst index.

In comparison to preceding models, our model incorporates a novel feature: it quantifies the contribution of each input parameter to the final forecast. Therefore, we conducted an analysis to determine the attribution of input parameters to the outcomes across different data, seeking to elucidate the underlying physical processes that drive the predictive performance of our machine-learning black box. In the temporal dimension, it is evident that the more recent the time, the more substantial the influence on the final prediction, and those data earlier have a very low contribution to the Dst-index prediction result. However, for the geomagnetic-storm period, the data from the last 3 – 9 h contribute more to the forecast relative to the quiet period, which means the earlier observation is essential in the space-weather forecast. Turning to the physical parameters, the historical Dst index emerges as the most influential factor in the final prediction results. In addition, the z-component of the magnetic field and the proton density also have essential contributions to the final prediction. Notably, as the Dst index decreases, the contribution of the z-component of the magnetic field and the

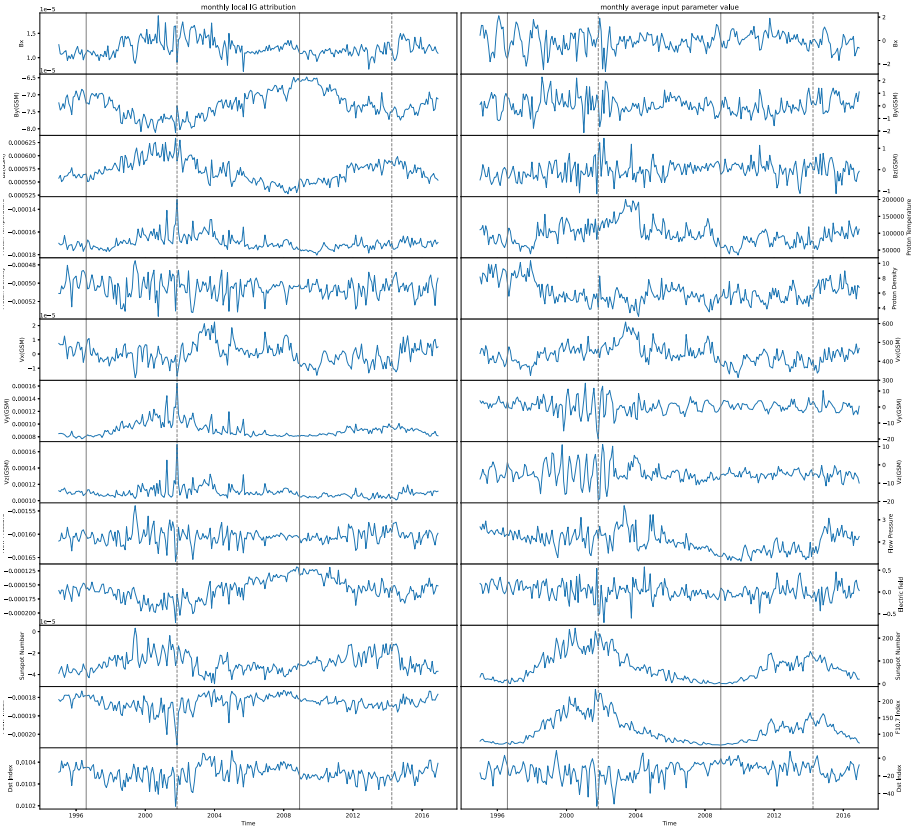

Figure 20 The left figure depicts the relationship between the local attribution and observation time for each input parameter. The right figure illustrates the average input values for various input parameters over time. Within these two figures, graphs at the same vertical heights represent the same physical input parameter. The vertical solid lines denote the solar minimum, while the vertical dashed lines represent the solar maximum.

y-component of the solar-wind velocity increase rapidly, which means they have a strong relationship with the geomagnetic storms.

In addition to analyzing the influence of various interplanetary sources of geomagnetic storms, the corresponding changes during different geomagnetic-storm periods are also conducted. Apart from the analogous changes for the geomagnetic-storm periods, the ICME and SIR periods also show deviations. The proton temperature and the sunspot number are the main parameters that have different attribution during ICME and SIR periods. Furthermore, we investigate variations across different solar cycles and years, yielding significant findings. It is observed that certain input variables, particularly the sunspot number, exhibit varying impacts on the final prediction under different solar-activity levels and cycles. These results learned by machine are roughly consistent with the previous analytical models, underlining the utility and trustworthiness of our Dst-index prediction model. Nonetheless, comprehending the intricate workings of the machine-learning model remains a challenging task. Consequently, the quest for a more simplified model and a more interpretable AI algorithm is imperative to achieve a deeper understanding of the exact physical processes with the aid of artificial intelligence.

Although we can accurately predict the Dst index for the near future, forecasting the Dst index over an extended period remains challenging. One contributing factor is that the interplanetary data are mostly observed at the Sun–Earth Lagrangian point 1 (L1). It takes only several hours for the solar wind to travel from the Lagrangian point to Earth. Therefore, if we want to make accurate predictions longer in advance, the observation from Lagrangian point 1 (L1) of the Sun–Earth system is insufficient. To enhance the accuracy of space-weather forecasts within Earth's orbit, it is advisable to incorporate observation data from regions closer to the Sun, such as the Venus orbit (Shen et al. 2021). Some recently launched satellites such as the Parker Solar Probe (Fox et al. 2016) and the Solar Orbiter (Müller et al. 2020) are also observed in closer orbits, which may further improve the prediction accuracy. Considering that the Sun is rotating, the observation data not far to the east of the Sun–Earth line can also be used for earlier space-weather forecasts (Chi et al. 2022). The two STEREO satellites (Kaiser et al. 2008) pass through this area at different times and the solar-ring mission (Wang et al. 2023) plans to place a satellite in this area in the near future.

Author contributions Junyan Liu and Chenglong Shen wrote the main manuscript. Yang Wang, Mengjiao Xu, Yutian Chi, Zhihui Zhong, Dongwei Mao, Can Wang, Jiajia Liu, and Yuming Wang contributed to the discussions and offered revision suggestions.

Funding This work is supported by grants from the NSFC (42325405, 42130204, 41904151, 42188101, 42074222) and the Strategic Priority Program of the Chinese Academy of Sciences (XDB41000000).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Abduallah, Y., Wang, J.T.L., Bose, P., Zhang, G., Gerges, F., Wang, H.: 2022, Forecasting the disturbance storm time index with bayesian deep learning. *Int. FLAIRS Conf. Proc.* 35. DOI.

Akasofu, S.-I.: 1981, Prediction of development of geomagnetic storms using the solar wind-magnetosphere energy coupling function ϵ. *Planet. Space Sci.* **29**, 1151. DOI.

Ancona, M., Ceolini, E., Öztireli, C., Gross, M.: 2017, Towards Better Understanding of Gradient-Based Attribution Methods for Deep Neural Networks. ArXiv. DOI.

Asensio Ramos, A., Cheung, M.C.M., Chifu, I., Gafeira, R.: 2023, Machine learning in solar physics. *Living Rev. Solar Phys.* **20**, 4. DOI.

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., Mueller, K.-R.: 2009, How to Explain Individual Classification Decisions. ArXiv.

Bai, S., Kolter, J.Z., Koltun, V.: 2018, An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv [cs].

Boynton, R.J., Balikhin, M.A., Billings, S.A., Sharma, A.S., Amariutei, O.A.: 2011, Data derived NARMAX Dst model. *Ann. Geophys.* **29**, 965. DOI.

Branco, P., Torgo, L., Ribeiro, R.: 2015, A Survey of Predictive Modelling under Imbalanced Distributions. ArXiv.

Buhrmester, V., Münch, D., Arens, M.: 2021, Analysis of explainers of black box deep neural networks for computer vision: a survey. *Mach. Learn. Knowl. Extr.* **3**, 966. DOI.

Burton, R.K., McPherron, R.L., Russell, C.T.: 1975, An empirical relationship between interplanetary conditions and Dst. *J. Geophys. Res.* **80**, 4204. DOI.

Cai, L., Ma, S.Y., Zhou, Y.L.: 2010, Prediction of SYM-H index during large storms by NARX neural network from IMF and solar wind data. *Ann. Geophys.* **28**, 381. DOI.

Chi, Y., Shen, C., Wang, Y., Xu, M., Ye, P., Wang, S.: 2016, Statistical study of the interplanetary coronal mass ejections from 1995 to 2015. *Solar Phys.* **291**, 2419. DOI.

Chi, Y., Shen, C., Luo, B., Wang, Y., Xu, M.: 2018, Geoeffectiveness of stream interaction regions from 1995 to 2016. *Adv. Space Res.* **16**, 1960. DOI.

Chi, Y., Shen, C., Scott, C., Xu, M., Owens, M., Wang, Y., Lockwood, M.: 2022, Predictive capabilities of corotating interaction regions using STEREO and wind in-situ observations. *Adv. Space Res.* **20**, e2022SW003112. DOI.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: 2014, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv [cs].

Cliver, E.W., Kamide, Y., Ling, A.G.: 2000, Mountains versus valleys: semiannual variation of geomagnetic activity. *J. Geophys. Res. Space Phys.* **105**, 2413. DOI.

Daly, E.J.: 2005, Outlook on space weather effects on spacecraft. In: Daglis, I.A. (ed.) *Effects of Space Weather on Technology Infrastructure* **91**, Springer, Dordrecht, 978.

Dang, T., Li, X., Luo, B., Li, R., Zhang, B., Pham, K., Ren, D., Chen, X., Lei, J., Wang, Y.: 2022, Unveiling the space weather during the starlink satellites destruction event on 4 February 2022. *Adv. Space Res.* **20**, e2022SW003152. DOI.

Echer, E., Gonzalez, W.D., Alves, M.V.: 2006, On the geomagnetic effects of solar wind interplanetary magnetic structures. *Adv. Space Res.* **4**, 2005SW000200. DOI.

Elman, J.L.: 1990, Finding structure in time. *Cogn. Sci.* **14**, 179. DOI.

Fox, N.J., Velli, M.C., Bale, S.D., Decker, R., Driesman, A., Howard, R.A., Kasper, J.C., Kinnison, J., Kusterer, M., Lario, D., Lockwood, M.K., McComas, D.J., Raouafi, N.E., Szabo, A.: 2016, The solar probe plus mission: humanity's first visit to our star. *Space Sci. Rev.* **204**, 7. DOI.

Gleisner, H., Lundstedt, H., Wintoft, P.: 1996, Predicting geomagnetic storms from solar-wind data using time-delay neural networks. *Ann. Geophys.* **14**, 679. DOI.

Gonzalez, W.D., Echer, E.: 2005, A study on the peak dst and peak negative bz relationship during intense geomagnetic storms. *Geophys. Res. Lett.* **32**, 2005GL023486. DOI.

Gonzalez, W.D., Joselyn, J.A., Kamide, Y., Kroehl, H.W., Rostoker, G., Tsurutani, B.T., Vasyliunas, V.M.: 1994, What is a geomagnetic storm? *J. Geophys. Res. Space Phys.* **99**, 5771. DOI.

Gopalswamy, N., Barbieri, L., Cliver, E.W., Lu, G., Plunkett, S.P., Skoug, R.M.: 2005, Introduction to violent Sun-Earth connection events of October–November 2003. *J. Geophys. Res. Space Phys.* **110**, 2005JA011268. DOI.

Gruet, M.A., Chandorkar, M., Sicard, A., Camporeale, E.: 2018, Multiple-hour-ahead forecast of the dst index using a combination of long short-term memory neural network and Gaussian process. *Adv. Space Res.* **16**, 1882. DOI.

Hapgood, M., Liu, H., Lugaz, N.: 2022, SpaceX—sailing close to the space weather? *Adv. Space Res.* **20**, e2022SW003074. DOI.

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M.H., Brett, M., Haldane, A., Del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: 2020, Array programming with NumPy. *Nature* **585**, 357. DOI.

He, H., Garcia, E.A.: 2009, Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**, 1263. DOI.

He, K., Zhang, X., Ren, S., Sun, J.: 2016, Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **770**, 1063. DOI.

Heirtzler, J.R., Allen, J.H., Wilkinson, D.C.: 2002, Ever-present South Atlantic anomaly damages spacecraft. *Eos Trans. AGU* **83**, 165. DOI.

Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., Liu, Y.: 2020, Temporal Convolutional Neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Comput.* **24**, 16453. DOI.

Hochreiter, S., Schmidhuber, J.: 1997, Long short-term memory. *Neural Comput.* **9**, 1735. DOI.

Hu, A., Camporeale, E., Swiger, B.: 2023, Multi-hour-ahead Dst index prediction using multi-fidelity boosted neural networks. *Adv. Space Res.* **21**, e2022SW003286. DOI.

Hunter, J.D.: 2007, Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90. DOI.

Hutter, F., Hoos, H.H., Leyton-Brown, K.: 2011, Sequential model-based optimization for general algorithm configuration. In: Coello, C.A.C. (ed.) *Learning and Intelligent Optimization*, *Lecture Notes in Computer Science* **6683**, Springer, Berlin, 978. DOI.

Kaiser, M.L., Kucera, T.A., Davila, J.M., St.Cyr, O.C., Guhathakurta, M., Christian, E.: 2008, The STEREO mission: an introduction. *Space Sci. Rev.* **136**, 5. DOI.

Kappenman, J.G.: 2005, An overview of the impulsive geomagnetic field disturbances and power grid impacts associated with the violent sun-Earth connection events of 29-31 October 2003 and a comparative evaluation with other contemporary storms. *Adv. Space Res.* **3**, 2004SW000128. DOI.

King, J.H., Papitashvili, N.E.: 2005, Solar wind spatial scales in and comparisons of hourly wind and ACE plasma and magnetic field data. *J. Geophys. Res. Space Phys.* **110**, 2004JA010649. DOI.

Krawczyk, B.: 2016, *Learning from Imbalanced Data: Open Challenges and Future Directions*, *Lect. Notes. Artif. Int.* **5**, 221. DOI.

Lazzús, J.A., Vega, P., Rojas, P., Salfate, I.: 2017, Forecasting the Dst index using a swarm-optimized neural network: forecasting the Dst index using ANN+PSO. *Adv. Space Res.* **15**, 1068. DOI.

Lethy, A., El-Eraki, M.A., Samy, A., Deebes, H.A.: 2018, Prediction of the Dst index and analysis of its dependence on solar wind parameters using neural network. *Adv. Space Res.* **16**, 1277. DOI.

Li, Y.Y., Huang, S.Y., Xu, S.B., Yuan, Z.G., Jiang, K., Wei, Y.Y., Zhang, J., Xiong, Q.Y., Wang, Z., Lin, R.T., Yu, L.: 2022, Selection of the main control parameters for the Dst index prediction model based on a layer-wise relevance propagation method. *Astron. Astrophys. Suppl. Ser.* **260**, 6. DOI.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: 2020, Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 318. DOI.

Lindauer, M., Eggensperger, K., Feurer, M., Biedenkapp, A., Deng, D., Benjamins, C., Ruhkopf, T., Sass, R., Hutter, F.: 2022, SMAC3: a versatile Bayesian optimization package for hyperparameter optimization. *J. Mach. Learn. Res.* **23**, 1.

Lühr, H., Maus, S.: 2010, Solar cycle dependence of quiet-time magnetospheric currents and a model of their near-Earth magnetic fields. *Earth Planets Space* **62**, 843. DOI.

Lundstedt, H., Wintoft, P.: 1994, Prediction of geomagnetic storms from solar wind data with the use of a neural network. *Ann. Geophys.* **12**, 19. DOI.

Lundstrom, D.D., Huang, T., Razaviyayn, M.: 2022, A rigorous study of integrated gradients method and extensions to internal neuron attributions. In: Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., Sabato, S. (eds.) *Proceedings of the 39th International Conference on Machine Learning*, *Proceedings of Machine Learning Research* **162**, PMLR, 14485.

Mandea, M., Chambodut, A.: 2020, Geomagnetic field processes and their implications for space weather. *Surv. Geophys.* **41**, 1611. DOI.

Mannucci, A.J., Tsurutani, B.T., Iijima, B.A., Komjathy, A., Saito, A., Gonzalez, W.D., Guarnieri, F.L., Kozyra, J.U., Skoug, R.: 2005, Dayside global ionospheric response to the major interplanetary events of October 29–30, 2003 "halloween storms". *Geophys. Res. Lett.* **32**, 2004GL021467. DOI.

Microsoft: 2021, Neural Network Intelligence.

Molinski, T.S., Feero, W.E., Damsky, B.L.: 2000, Shielding grids from solar storms [power system protection]. *IEEE Spectr.* **37**, 55. DOI.

Müller, D., St.Cyr, O.C., Zouganelis, I., Gilbert, H.R., Marsden, R., Nieves-Chinchilla, T., Antonucci, E., Auchère, F., Berghmans, D., Horbury, T.S., Howard, R.A., Krucker, S., Maksimovic, M., Owen, C.J., Rochus, P., Rodriguez-Pacheco, J., Romoli, M., Solanki, S.K., Bruno, R., Carlsson, M., Fludra, A., Harra, L., Hassler, D.M., Livi, S., Louarn, P., Peter, H., Schühle, U., Teriaca, L., Del Toro Iniesta, J.C., Wimmer-Schweingruber, R.F., Marsch, E., Velli, M., De Groof, A., Walsh, A., Williams, D.: 2020, The solar orbiter mission – science overview. *Astron. Astrophys.* **642**, A1. DOI.

Mursula, K., Karinen, A.: 2005, Explaining and correcting the excessive semiannual variation in the Dst index. *Geophys. Res. Lett.* **32**, 2005GL023132. DOI.

Nose, M., Sugiura, M., Kamei, T., Iyemori, T., Koyama, Y.: 2015, *Dst Index*, WDC for Geomagnetism, Kyoto. DOI.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: 2019, PyTorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., dAlché-Buc, F., Fox, E., Garnett, R. (eds.) *Advances in Neural Information Processing Systems* **32**, Curran Associates, Red Hook, 8024.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: 2011, Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825.

Perreault, P., Akasofu, S.-I.: 1978, A study of geomagnetic storms. *Geophys. J. Int.* **54**, 547. DOI.

Pulkkinen, A., Bernabeu, E., Eichner, J., Beggan, C., Thomson, A.W.P.: 2012, Generation of 100-year geomagnetically induced current scenarios. *Adv. Space Res.* **10**, 2011SW000750. DOI.

Russell, C.T., McPherron, R.L., Burton, R.K.: 1974, On the cause of geomagnetic storms. *J. Geophys. Res.* **79**, 1105. DOI.

Schulte In Den Bäumen, H., Moran, D., Lenzen, M., Cairns, I., Steenge, A.: 2014, How severe space weather can disrupt global supply chains. *Nat. Hazards Earth Syst. Sci.* **14**, 2749. DOI.

Shen, C., Chi, Y., Wang, Y., Xu, M., Wang, S.: 2017, Statistical comparison of the ICME's geoeffectiveness of different types and different solar phases from 1995 to 2014. *J. Geophys. Res. Space Phys.* **122**, 5931. DOI.

Shen, C., Chi, Y., Xu, M., Wang, Y.: 2021, *Origin of Extremely Intense Southward Component of Magnetic Field (Bs) in ICMEs*, *Aip. Conf. Proc.* **9**, 762488. DOI.

Shrikumar, A., Greenside, P., Kundaje, A.: 2019, Learning Important Features Through Propagating Activation Differences. arXiv [cs].
Simonyan, K., Vedaldi, A., Zisserman, A.: 2014, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. ArXiv.
Sugiura, M.: 1963, Hourly Values of Equatorial Dst for the Igy. Technical Report NASA-TM-X-55238.
Sugria, M., Chapman, S.: 1961, The average morphology of geomagnetic storms with sudden commencement. *Abhandl, Akad, Wiss, Goettingen Math, Phys., Kl* **4**, 1.
Sundararajan, M., Taly, A., Yan, Q.: 2017, Axiomatic Attribution for Deep Networks. arXiv [cs].
Temerin, M., Li, X.: 2015, The dst index underestimates the solar cycle variation of geomagnetic activity. *J. Geophys. Res. Space Phys.* **120**, 5603. DOI.
The pandas development team: 2020, Pandas-Dev/Pandas: Pandas, Zenodo. DOI.
Tsurutani, B.T., Gonzalez, W.D., Kamide, Y., Arballo, J.K.: 1997, Preface. In: Tsurutani, B.T., Gonzalez, W.D., Kamide, Y., Arballo, J.K. (eds.) *Geophysical Monograph Series* **98**, American Geophysical Union, Washington, 978. DOI.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Van Mulbregt, P., SciPy 1.0 Contributors, Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., De Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y. (eds.): 2020, SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* **17**, 261. DOI.
Wang, Y., Shen, C.L., Wang, S., Ye, P.Z.: 2003, An empirical formula relating the geomagnetic storm's intensity to the interplanetary parameters: – and Δ t. *Geophys. Res. Lett.* **30**, 2003GL017901. DOI.
Wang, Y., Ye, P., Zhou, G., Wang, S., Wang, S., Yan, Y., Wang, J.: 2005, The interplanetary responses to the great solar activities in late October 2003. *Solar Phys.* **226**, 337. DOI.
Wang, Y., Bai, X., Chen, C., Chen, L., Cheng, X., Deng, L., Deng, L., Deng, Y., Feng, L., Gou, T., Guo, J., Guo, Y., Hao, X., He, J., Hou, J., Huang, J., Huang, Z., Ji, H., Jiang, C., Jiang, J., Jin, C., Li, X., Li, Y., Liu, J., Liu, K., Liu, L., Liu, R., Liu, R., Qiu, C., Shen, C., Shen, F., Shen, Y., Shi, X., Su, J., Su, Y., Su, Y., Sun, M., Tan, B., Tian, H., Wang, Y., Xia, L., Xie, J., Xiong, M., Xu, M., Yan, X., Yan, Y., Yang, S., Yang, S., Zhang, S., Zhang, Q., Zhang, Y., Zhao, J., Zhou, G., Zou, H.: 2023, Solar ring mission: building a panorama of the sun and inner-heliosphere. *Adv. Space Res.* **71**, 1146. DOI.
Wei, H.L., Zhu, D.Q., Billings, S.A., Balikhin, M.A.: 2007, Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks. *Adv. Space Res.* **40**, 1863. DOI.
Wolpert, D.H.: 1992, Stacked generalization. *Neural Netw.* **5**, 241. DOI.
Wu, J.-G., Lundstedt, H.: 1997, Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks. *J. Geophys. Res. Space Phys.* **102**, 14255. DOI.
Xu, S.B., Huang, S.Y., Yuan, Z.G., Deng, X.H., Jiang, K.: 2020, Prediction of the Dst index with bagging ensemble-learning algorithm. *Astrophys. J. Suppl.* **248**, 14. DOI.
Yang, Y., Zha, K., Chen, Y.-C., Wang, H., Katabi, D.: 2021, Delving into deep imbalanced regression. In: *International Conference on Machine Learning (ICML)*.
Yang, Y., Liu, J.J., Feng, X.S., Chen, P.F., Zhang, B.: 2023, Prediction of the transit time of coronal mass ejections with an ensemble machine-learning method. *Astrophys. J. Suppl.* **268**, 69. DOI.
Yermolaev, Y.I., Yermolaev, M.Y.: 2002, Statistical relationships between solar, interplanetary, and geomagnetospheric disturbances, 1976–2000. *Cosm. Res.* **40**, 1. DOI.
Zhang, J., Feng, Y., Zhang, J., Li, Y.: 2023, The short time prediction of the dst index based on the long-short time memory and empirical mode decomposition–long-short time memory models. *Appl. Sci.* **13**, 11824. DOI.
Zhao, H., Zong, Q., Wei, Y., Wang, Y.: 2011, Influence of solar wind dynamic pressure on geomagnetic Dst index during various magnetic storms. *Sci. China, Technol. Sci.* **54**, 1445. DOI.

Authors and Affiliations

Junyan Liu[1] · Chenglong Shen[2,3] · Yang Wang[4,5] · Mengjiao Xu[6] · Yutian Chi[6] · Zhihui Zhong[1] · Dongwei Mao[1] · Zhiyong Zhang[1] · Can Wang[1] · Jiajia Liu[2,3] · Yuming Wang[2,3,7]

✉ J. Liu
liujunyan@mail.ustc.edu.cn

✉ C. Shen
clshen@ustc.edu.cn

1 CAS Key Laboratory of Geospace Environment, Department of Geophysics and Planetary Sciences, University of Science and Technology of China, Hefei, China

2 Deep Space Exploration Laboratory/School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China

3 CAS Center for Excellence in Comparative Planetology, University of Science and Technology of China, Hefei, China

4 School of Software Engineering, University of Science and Technology of China, Hefei 230026, China

5 School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China

6 Institute of Deep Space Sciences, Deep Space Exploration Laboratory, Hefei 230026, China

7 Anhui Mengcheng Geophysics National Observation and Research Station, University of Science and Technology of China, Mengcheng, Anhui, China