Solar Physics
DOI: 10.1007/●●●●●-●●●-●●●-●●●●-●

Prediction of Geoeffective CMEs Using SOHO Images and Deep Learning

Khalid A. Alobaid[1,2] · **Jason T. L. Wang**[1,3] ·
Haimin Wang[1,4,5] · **Ju Jing**[1,4,5] ·
Yasser Abduallah[1,3] · **Zhenduo Wang**[1,3] ·
Hameedullah Farooki[6] · **Huseyin Cavus**[7] ·
Vasyl Yurchyshyn[5]

Abstract The application of machine learning to the study of coronal mass ejections (CMEs) and their impacts on Earth has seen significant growth recently. Understanding and forecasting CME geoeffectiveness is crucial for protecting infrastructure in space and ensuring the resilience of technological systems on Earth. Here we present GeoCME, a deep-learning framework designed to predict, deterministically or probabilistically, whether a CME event that arrives at Earth will cause a geomagnetic storm. A geomagnetic storm is defined as a disturbance of the Earth's magnetosphere during which the minimum Dst index value is less than -50 nT. GeoCME is trained on observations from the instruments including

K. A. Alobaid
kalobaid@ksu.edu.sa

J. T. L. Wang
wangj@njit.edu

H. Wang
haimin.wang@njit.edu

1 Institute for Space Weather Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

2 College of Applied Computer Sciences, King Saud University, Riyadh 11451, Saudi Arabia

3 Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

4 Center for Solar-Terrestrial Research, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

5 Big Bear Solar Observatory, New Jersey Institute of Technology, 40386 North Shore Lane, Big Bear City, CA 92314, USA

6 Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

7 Department of Physics, Canakkale Onsekiz Mart University, 17110 Canakkale, Turkey

LASCO C2, EIT and MDI on board the Solar and Heliospheric Observatory (SOHO), focusing on a dataset that includes 136 halo/partial halo CMEs in Solar Cycle 23. Using ensemble and transfer learning techniques, GeoCME is capable of extracting features hidden in the SOHO observations and making predictions based on the learned features. Our experimental results demonstrate the good performance of GeoCME, achieving a Matthew's correlation coefficient of 0.807 and a true skill statistics score of 0.714 when the tool is used as a deterministic prediction model. When the tool is used as a probabilistic forecasting model, it achieves a Brier score of 0.094 and a Brier skill score of 0.493. These results are promising, showing that the proposed GeoCME can help enhance our understanding of CME-triggered solar-terrestrial interactions.

Keywords: Coronal mass ejections; Solar-terrestrial relations; Heliosphere

1. Introduction

The impacts of geomagnetic storms on Earth have been investigated by many researchers (e.g., Wanliss and Showalter, 2006; Newell et al., 2007; Baker et al., 2013; Schrijver et al., 2015; Augusto et al., 2018; Joshi et al., 2018; Haines et al., 2019; Wu et al., 2019; Abunin et al., 2020; Chertok, 2020; Mishra et al., 2021; Besliu-Ionescu, Maris Muntean, and Dobrica, 2022; Pal, Nandy, and Kilpua, 2022; Raghav et al., 2023; Zhang et al., 2023; Hayakawa, Ebihara, and Pevtsov, 2024; Melkumyan et al., 2024). These storms can affect the accuracy of technological systems, such as satellites and communication systems, that rely on precise measurements of the Earth's magnetic field. They can also affect power grids by inducing electrical currents that can damage or disrupt the operation of the grids. In general, geomagnetic storms occur due to the interaction between radiation and plasma released by the Sun into the heliosphere and magnetic fields in the plasma environment near Earth (Wanliss and Showalter, 2006). The degree of severity exhibited by a storm is assessed through geomagnetic indices, such as the Kp index (Planetary K-index), the AE (Auroral Electrojet) index, and the Dst (Disturbance Storm Time) index. Mayaud (1980) discussed the meaning of these indices. Other geomagnetic indices, such as the SYM-H index and the ASY-H index, are similar to the Dst index, but are available in high resolution, with intervals as short as 1 minute or 5 minutes (Wanliss and Showalter, 2006).

Coronal mass ejections (CMEs), which carry strong southward-directed magnetic fields, may cause intense geomagnetic storms (Vourlidas, Patsourakos, and Savani, 2019; Baratashvili et al., 2022; Martinić et al., 2023). Predicting whether a CME will hit Earth and when it will reach Earth is a challenging task. Efforts to tackle this task include the use of empirical models (e.g., Brueckner et al., 1998; Manoharan et al., 2004; Gopalswamy et al., 2005), drag-based models (e.g., Vršnak and Gopalswamy, 2002; Dumbović et al., 2021), physics-based models (e.g., Fry et al., 2001; Moon et al., 2002) and machine learning models (e.g., Liu et al., 2018; Alobaid et al., 2022; Guastavino et al., 2023; Yang et al., 2023; Chierichini et al., 2024), among others (e.g., Zhao and Dryer, 2014; Singh et al., 2023). Machine learning models have also been used to predict the geoeffectiveness of CMEs. For example, Besliu-Ionescu et al. (2019) adopted logistic regression with numerical CME parameters to make predictions. Fu et al. (2021) presented a deep neural network to predict the geoeffectiveness and arrival time of CMEs. The authors used data from SOHO's Large Angle and Spectrometric Coronagraph (LASCO) C2 Field-of-View (FOV) and Extreme Ultraviolet Imaging Telescope (EIT), along with SDO's Atmospheric Imaging Assembly (AIA) observations. Pricopi et al. (2022) explored several machine learning methods such as logistic regression, k-nearest neighbors, and support vector machines, together with solar onset parameters, to predict the geoeffectiveness of CMEs.

The aforementioned studies predict CMEs that reach Earth and cause geomagnetic storms as "geoeffective," and predict all others, including CMEs that do not reach Earth, as "non-geoeffective." In contrast, we focus on CMEs that arrive at Earth, and predict whether they will cause geomagnetic storms. Since the problem we attempt to solve here differs from those addressed in previous work, the way we collect data for model training and testing is different from those used in previous work. The criterion for a disturbance of the Earth's magnetosphere to be considered a geomagnetic storm is that its minimum Dst value must be less than -50 nT (Gonzalez et al., 1994; Telloni, 2022). The Dst index, measured in nanoteslas (nT), is a key indicator used in space weather research to quantify the intensity of geomagnetic storms (Mayaud, 1980). It reflects the effect of geomagnetic disturbances caused by solar activity on Earth, where lower Dst values correspond to stronger storms.

Our work is based on SOHO observations, including LASCO C2 and EIT images, as well as Michelson Doppler Imager (MDI) magnetograms. Our goal is to understand whether machine learning can capture any possible connection between the SOHO observations and CME geoeffectiveness. We propose a deep learning framework, named GeoCME, to achieve this goal. Our main assumption is that the CMEs at hand have already arrived at Earth. In practice, how do we know whether a CME can reach Earth? This question can be answered using existing CME arrival prediction methods (e.g., Sudar, Vršnak, and Dumbović, 2016; Liu et al., 2018; Amerstorfer et al., 2021; Dumbović et al., 2021; Kaportseva and Shugay, 2021; Baratashvili et al., 2022; Guastavino et al., 2023; Chierichini et al., 2024). Thus, the use of GeoCME is a two-step process. In the first step, we use the existing methods mentioned above to predict whether a CME would arrive at Earth. If the CME is predicted to reach Earth, then in the second step we use GeoCME to predict whether the CME will cause a geomagnetic storm, i.e., whether the CME is geoeffective.

The remainder of this paper is organized as follows. Section 2 describes the data used in our study. Section 3 presents the architecture and configuration details of GeoCME. Section 4 reports the experimental results. Section 5 presents a discussion and concludes the article.

2. Data

We focused on halo/partial halo CMEs in Solar Cycle 23 (Michalek et al., 2006; Gopalswamy, Yashiro, and Akiyama, 2007; Gopalswamy, 2009). Figure

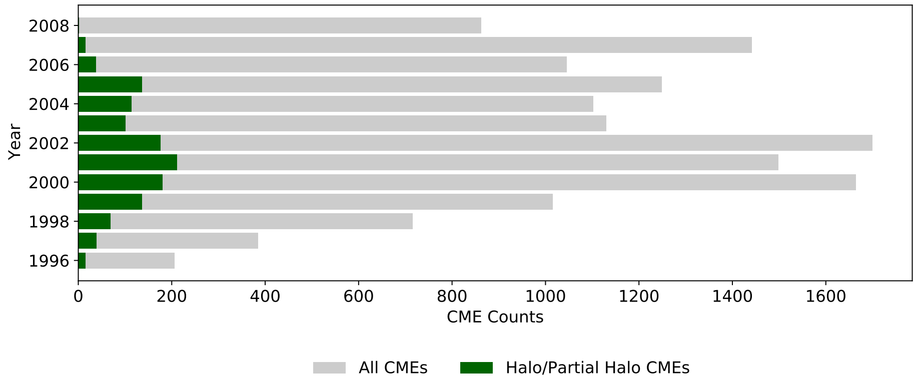

Figure 1. Chart showing the total counts of halo/partial halo CMEs among all CMEs during Solar Cycle 23 (1996-2008) according to the SOHO/LASCO CME catalog.

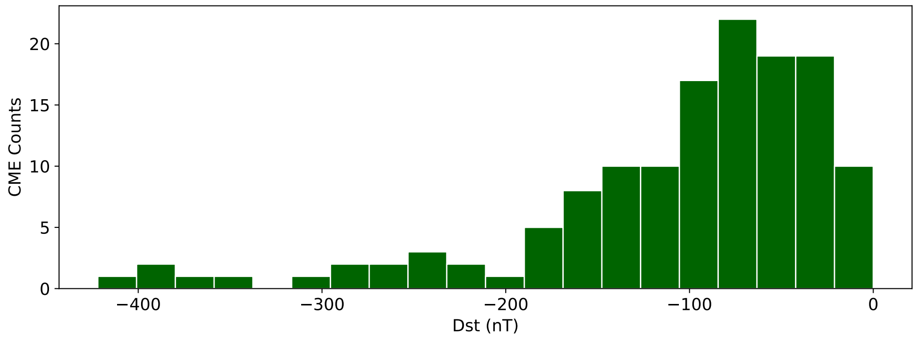

Figure 2. Distribution of the Dst index values caused by the 136 halo/partial halo CME events in our dataset.

1 shows the total counts of halo/partial halo CMEs in Cycle 23 according to the SOHO/LASCO CME catalog (Yashiro et al., 2004). The CME events used in our study were obtained from the list of interplanetary coronal mass ejections (ICMEs), known as the RC list, compiled and maintained by Richardson and Cane (2010). We chose 145 CME events within the RC list that occurred in Solar Cycle 23 and arrived at Earth (i.e. with arrival-time data). We used the SOHO/LASCO CME catalog to identify and select 141 halo/partial halo CME events among the 145 CME events. The RC list shows the minimum Dst index value caused by a CME during its interplanetary interaction with the Earth's magnetosphere. We excluded those CME events without Dst index values, which resulted in a total of 136 halo/partial-halo CME events. Figure 2 shows the distribution of the minimum Dst values caused by the 136 events. As mentioned in the previous section, a value of -50 nT was used for the Dst index to determine the geoeffectiveness of CMEs (Gonzalez et al., 1994; Telloni, 2022). As a consequence, among the 136 halo/partial halo CME events analyzed, 101 were identified as geoeffective, while 35 were classified as non-geoeffective. Figure 3 provides a breakdown analysis of the geoeffective and non-geoeffective

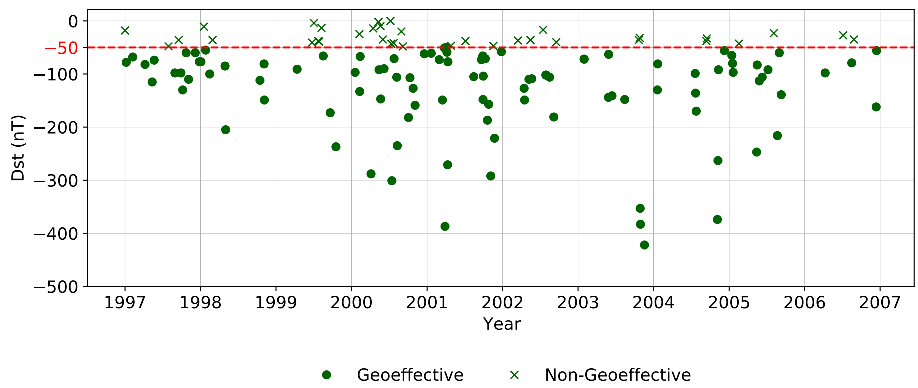

Figure 3. Breakdown analysis of the geoeffective and non-geoeffective CME events in our dataset where a solid circle represents a geoeffective CME event and a cross mark represents a non-geoeffective CME event. These events were distributed over 10 years, from 1997 to 2006.

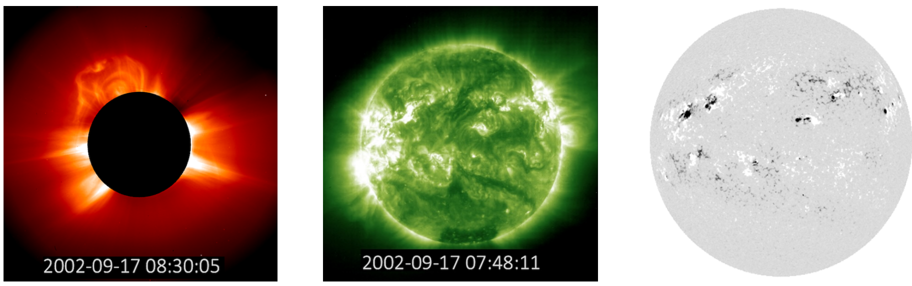

Figure 4. SOHO observations on the CME event that occurred at 08:06:00 UT on 17 September 2002. Shown from left to right are a LASCO C2 image, an EIT 195 Å image, and a full-disk MDI magnetogram.

CME events in our dataset. These events were distributed over 10 years, from 1997 to 2006.

When training and testing our GeoCME framework, we used three types of SOHO data, namely LASCO C2, EIT 195 Å, and MDI magnetogram images. Figure 4 shows the SOHO observations on the CME event that occurred at 08:06:00 UT on 17 September 2002, which are, from left to right, LASCO C2, EIT, and MDI, respectively. LASCO C2 coronagraph captures images of the Sun from 1.5 to 6 R$\odot$ (Brueckner et al., 1995). We constructed base-difference images for LASCO C2 by subtracting the pre-event image (base) from subsequent images of the event to enhance the visibility of dynamic solar features while minimizing static background information. This technique is widely used to improve machine learning of CME image features in LASCO C2 observations (Wang et al., 2019; Alobaid et al., 2023).

Our data collection process follows a systematic approach centered on CME appearance times, as listed in the SOHO/LASCO CME catalog. For LASCO C2, we collected images 10 minutes before a CME event and up to 4 hours after the event (Fu et al., 2021), averaging 12.34 images per event, totaling 1679 images. EIT images were collected from 4 hours before the event to the event time, with an average of 7.59 images per event, resulting in 1033 images. MDI magnetograms included the last three observations before the event, averaging 2.9 images per event for a total of 395 images. In total, this dataset contains 3107 images.

3. Methodology

3.1. Transfer Learning

We addressed the challenge of working with a relatively small dataset with 3107 images by using transfer learning. The transfer learning approach involved evaluating the efficacy of several pre-trained deep learning models, including ResNet (He et al., 2016), InceptionNet (Szegedy et al., 2016, 2017), VGG (Simonyan and Zisserman, 2015), MobileNet (Sandler et al., 2018), DenseNet (Huang et al., 2017), Xception (Chollet, 2017), and EfficientNet (Tan and Le, 2019). These pre-trained models were originally designed to perform representation learning, feature extraction, and image classification. Our experiments revealed that ResNet152 and InceptionResNetV2, both pre-trained on the ImageNet dataset (Deng et al., 2009), achieved the best results in geoeffective CME prediction.

Residual blocks in ResNet152 and inception modules in InceptionResNetV2 are core components that improve the performance of convolutional neural networks. Residual blocks help train very deep networks by allowing gradients to flow more easily through shortcut connections, thus solving the vanishing-gradient problem. Inception modules, on the other hand, use parallel convolutional filters of different sizes to capture image features at multiple scales. InceptionResNet combines residual blocks and inception modules, integrating residual connections with the inception structure to leverage the strengths of both. Figure 5 shows a residual block (He et al., 2016) in ResNet152 and an InceptionResNet module (Szegedy et al., 2016) in InceptionResNetV2. These components improve the accuracy and efficiency of deep neural networks, making them suitable for complex tasks such as recognizing patterns of solar imagery.

Our transfer learning approach, where a pre-trained image classification model is adapted to a new task (i.e., geoeffective CME prediction), provides an effective way to build a new model to specific needs without the substantial training data usually required for complex deep learning models.

3.2. The Ensemble Model

To further improve feature extraction capabilities, we combined ResNet152 and InceptionResNetV2, referred to as base models, into an integrated framework (GeoCME). This ensemble approach aimed to capitalize on the strengths of each base model, thereby enhancing the overall performance of the feature extraction process and, subsequently, the accuracy of geoeffective CME prediction. Figure 6 illustrates the architecture of the GeoCME framework, and Table 1 presents its configuration details. Table 2 summarizes the parameters of the base models used in GeoCME.

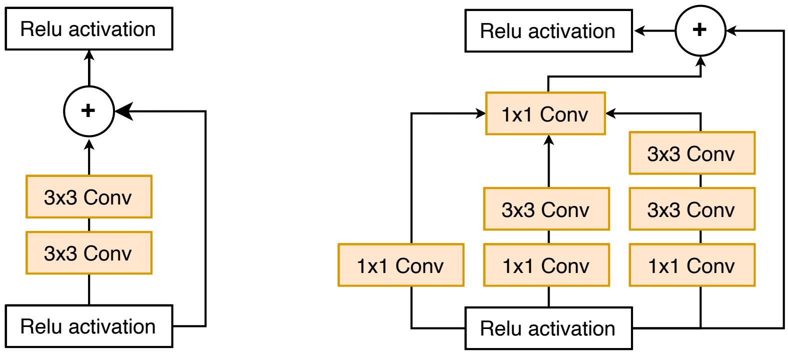

Figure 5. Illustration of a residual block (left) and an InceptionResNet module (right). The residual block consists of two 3×3 convolutional layers followed by a residual connection that allows gradients to flow directly through the network, improving training efficiency. The InceptionResNet module includes parallel convolutional paths with 1×1 and 3×3 filters, which, combined with a residual connection, capture image features at multiple scales to maintain efficient gradient flow.

For a given CME event, we feed the event's image from each instrument (LASCO C2, EIT, MDI) into the two base models, ResNet152 (RN) and InceptionResNetV2 (IRN), respectively. For each instrument, the output values of the two base models are fed into a concatenation layer. The concatenated features pass through three convolutional blocks (ConvBlock), each equipped with a 2D convolution layer with 64, 128, and 256 filters, respectively. Each convolutional block has a kernel size of 3×3, paired with LeakyReLU activation and batch normalization for stability. The features are flattened and processed through a dense layer of 1024 neurons. To avoid overfitting, a dropout layer with a rate of 0.3 is placed after the dense layer of 1024 neurons. The dropout layer is followed by another dense layer with 1 neuron.

As shown in Figure 6, each SOHO instrument (LASCO C2, EIT, MDI) uses the pipeline described above to produce one prediction per image. For a CME event with multiple images from the same instrument, an ensemble layer calculates the mean of all output values of the images from the same instrument for the CME event. At this stage, we have one predicted value per instrument for the CME event. The final ensemble layer then calculates the mean of the three predicted values from the three instruments to get the final output for the CME event. This final output is the probability that the CME event will be geoeffective, i.e. the CME event will cause a geomagnetic storm. We implemented a threshold of 0.6 in the output layer to obtain a deterministic model. If the probability is greater than or equal to the threshold, then the GeoCME model predicts that the CME event is geoeffective; otherwise, the model predicts that the CME event is non-geoeffective.

We have an imbalanced dataset at hand, which contains a positive (or majority) class with 101 geoeffective CME events and a negative (or minority) class with 35 non-geoeffective CME events. We use the Weighted Binary Cross Entropy (WBCE) loss function to combat the imbalance issue within the dataset

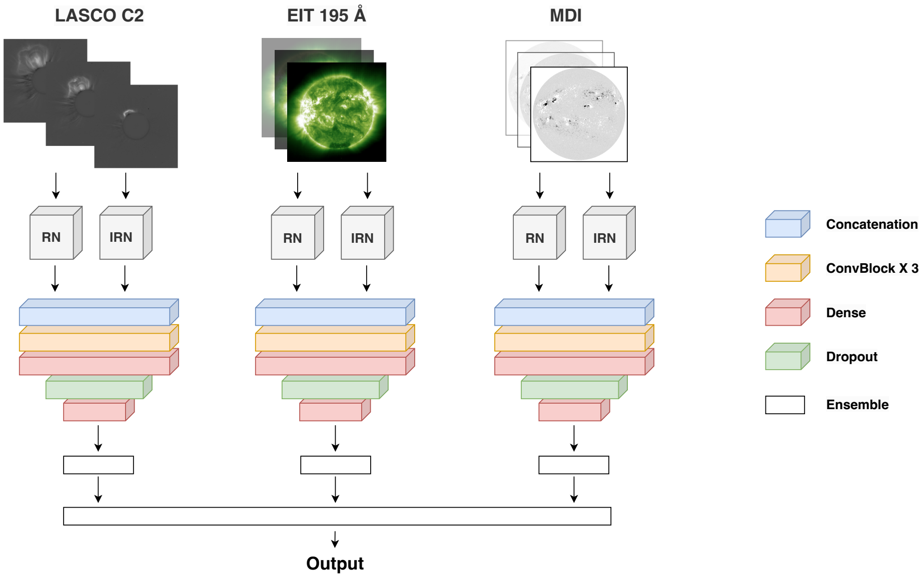

Figure 6. Illustration of the GeoCME architecture. The ensemble model consists of three equal pipelines (left, middle, right), each dedicated to one of the three SOHO instruments (LASCO C2, EIT, and MDI), respectively. Each pipeline begins with two base models, namely ResNet152 (RN) and InceptionResNetV2 (IRN), followed by a concatenation layer that combines the output values of the two base models. This concatenation layer is succeeded by three convolutional blocks, followed by two dense layers with 1024 neurons and 1 neuron, respectively, with a dropout layer between them. Each pipeline ends with an ensemble layer that produces the output of the corresponding SOHO instrument. Finally, the output values of the three pipelines corresponding to the three SOHO instruments are fed to another ensemble layer to produce the final result.

Table 1. Configuration details of the GeoCME framework.

Layer	Kernel No.	Kernel Size	Regularization	Activation	Output
ConvBlock 1	64	3×3	Batch Norm	LeakyReLU	$8 \times 8 \times 64$
ConvBlock 2	128	3×3	Batch Norm	LeakyReLU	$8 \times 8 \times 128$
ConvBlock 3	256	3×3	Batch Norm	LeakyReLU	$8 \times 8 \times 256$
Dense 1	–	–	Batch Norm	LeakyReLU	1024
Dense 2	–	–	–	Sigmoid	1

Table 2. Base model parameters.

Base Model	Layer Number	Parameter Number
ResNet152	152	58.50M
InceptionResNetV2	164	54.39M

Table 3. Hyperparameters for GeoCME training.

Loss Function	Optimizer	Dropout Rate	Batch Size	Epochs
WBCE	Adam	0.3	32	100

(Goodfellow, Bengio, and Courville, 2016; Liu et al., 2020; Abduallah et al., 2022). Let N denote the number of events in the training or validation set. Let w_0 denote the weight for the negative (or minority) class and let w_1 denote the weight for the positive (or majority) class. The weight assignment is based on the ratio of sizes between the majority and minority classes, with a higher weight assigned to the minority class, as shown below.

$$\mathrm{WBCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[w_0 y_i \log(\hat{y}_i) + w_1(1-y_i)\log(1-\hat{y}_i)\right]. \tag{1}$$

Here, y_i denotes the label of the ith event, with $y_i = 1$ for a geoeffective CME and $y_i = 0$ for a non-geoeffective CME, and $\hat{y}_i$ represents the predicted probability for the ith event being positive. The WBCE method ensures that the minority class is emphasized more in the loss calculation, effectively addressing the imbalance issue in our dataset. During model training, we use adaptive moment estimation (Adam) as optimizer (Goodfellow, Bengio, and Courville, 2016), with a batch size of 32, and a total of 100 epochs. Table 3 summarizes the hyperparameters used in model training. These hyperparameter values are obtained by using the grid search capability from the Python machine learning library, scikit-learn (Pedregosa et al., 2011). The validation set used for tuning the hyperparameters is described below.

4. Results

4.1. Experimental Setup

We adopted an 80:20 scheme to train and test the GeoCME framework. Specifically, we used 80% of the CME events from each of the "geoeffective" and "non-geoeffective" classes for model training and used the remaining 20% of the events from each class for model testing. Furthermore, we allocated 10% of the training data for each class for validation, so that the performance of our model was regularly evaluated against unseen data throughout the training process. Figure 7 presents the GeoCME training and validation learning curves. The downward and convergence trends in the learning curves demonstrate the effectiveness of GeoCME learning and its capacity to generalize successfully to new data. We note that the two base models of GeoCME (ResNet152 and InceptionResNetV2) are pre-trained on the extensive ImageNet dataset. We used a relatively small amount of new training data to retrain the two complex models for our use through transfer learning. Because the complex models have been

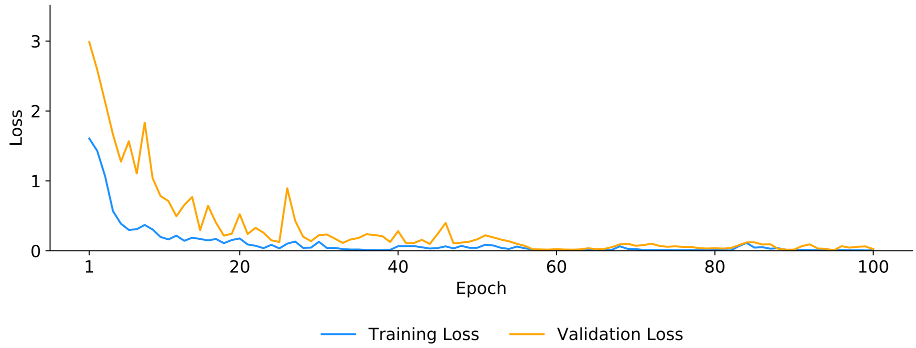

Figure 7. Training and validation learning curves showing GeoCME is a well-fit model for geoeffective CME prediction.

well pre-trained and GeoCME is a fusion of them, we see that the learning curves of GeoCME converge well in Figure 7.

In the experimental study, we adopted two metrics to evaluate GeoCME's performance: Matthew's Correlation Coefficient (MCC) and True Skill Statistics (TSS). Given a CME event E, we define E as a true positive (TP) if the model predicts E as positive (i.e. geoeffective) and E is indeed positive. We define E as a true negative (TN) if the model predicts E as negative (i.e. non-geoeffective) and E is indeed negative. We say that E is a false positive (FP) if the model predicts E as positive while E is actually negative; E is a false negative (FN) if the model predicts E as negative while E is actually positive. When the context is clear, we also use TP (TN, FP, and FN, respectively) to represent the total number of true positives (true negatives, false positives, and false negatives, respectively) produced by the model. The MCC and TSS are defined as follows (Liu et al., 2019; Abduallah et al., 2022):

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{2}$$

$$\mathrm{TSS} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} - \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}. \tag{3}$$

As mentioned above, we implemented a threshold in the GeoCME output layer to obtain a deterministic prediction model. If the probability produced by GeoCME for a given CME event is greater than or equal to the threshold, then the model predicts that the CME event is geoeffective; otherwise, the model predicts that the CME event is non-geoeffective. Figure 8 presents GeoCME's metric values for varying thresholds based on the validation set. The best metric values are obtained when the threshold is set to 0.6. As a consequence, we used the 0.6 threshold in our study.

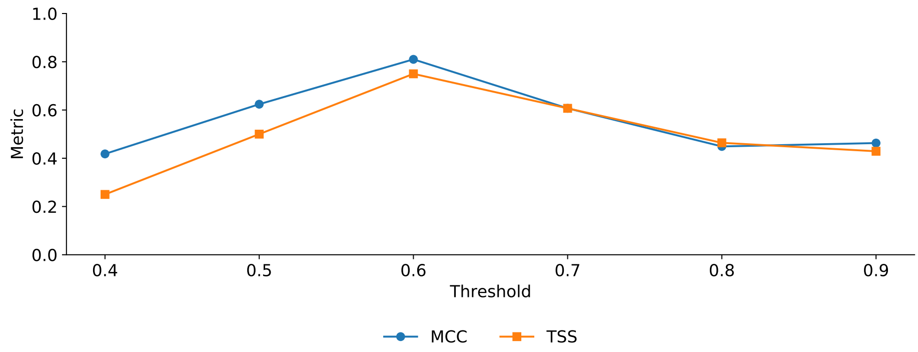

Figure 8. GeoCME's metric values for varying thresholds based on the validation set. The best metric values are obtained when the threshold is set to 0.6.

4.2. Performance Evaluation

We conducted ablation tests to analyze and evaluate the components of our GeoCME framework. GeoCME contains two pre-trained base models (see Figure 6): ResNet152 (RN) and InceptionResNetV2 (IRN). We considered three variants of GeoCME: GeoCME-RN-IRN, GeoCME-RN, GeoCME-IRN. GeoCME-RN-IRN denotes GeoCME with the RN and IRN models removed. This subnet contains only the inherent structure of GeoCME without the pre-trained models used for feature extraction. Thus, there is no transfer learning in GeoCME-RN-IRN. GeoCME-RN denotes GeoCME with the RN models removed. GeoCME-IRN denotes GeoCME with the IRN models removed. Figure 9 compares the performance of the four networks used as deterministic prediction models.

It can be seen in Figure 9 that the variants (GeoCME-RN, GeoCME-IRN, and GeoCME-RN-IRN), each missing a key component or more, achieved varied performance levels. The GeoCME framework, which integrates all its components, shows the best performance by achieving the highest MCC of 0.807 and the highest TSS of 0.714. GeoCME-RN-IRN, which lacks both the ResNet and InceptionResNet base models, exhibits the most significant drop in prediction accuracy, with the lowest MCC of 0.365 and the lowest TSS of 0.380. This highlights the impact of excluding transfer learning on GeoCME's performance. Furthermore, GeoCME-RN performs better than GeoCME-IRN, emphasizing the importance of InceptionResNet in improving the prediction accuracy.

Figure 10 presents the confusion matrix obtained by GeoCME, which provides a breakdown analysis of errors that occur when the model makes predictions in the test set. There are 28 CME events in the test set. Approximately (21+2)/28 = 82% of the events in the test set are predicted to be geoeffective. Approximately 2/(21+2) = 8.7% of the predictions are false alarms (false positives). The model's FP value is 2, indicating that it is a relatively sensitive model in the sense that it predicts two CME events as positive, while these events do not cause geomagnetic storms. However, the model does not miss any geomagnetic storms, as reflected by the fact that the model's FN value is zero.

Each CME event E is accompanied by images from three distinct SOHO instruments (LASCO C2, EIT and MDI). To evaluate the effectiveness of these im-

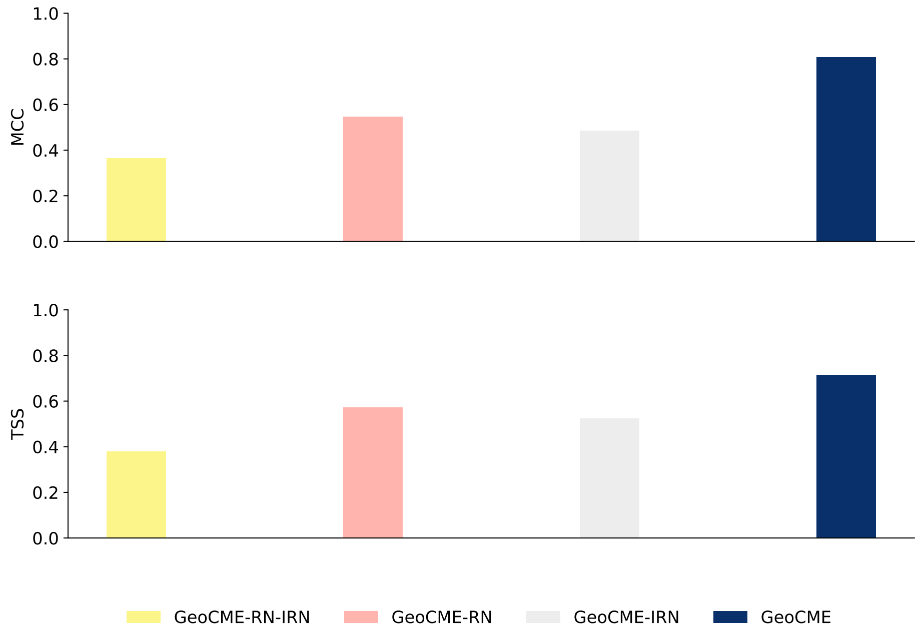

Figure 9. Results of the ablation tests for assessing four networks (GeoCME-RN-IRN, GeoCME-RN, GeoCME-IRN, and GeoCME) used as deterministic prediction models where GeoCME-RN-IRN represents GeoCME with RN and IRN models removed, GeoCME-RN represents GeoCME with RN models removed, GeoCME-IRN represents GeoCME with IRN models removed, and GeoCME represents the full model. (Top) MCC of the networks tested. (Bottom) TSS of the networks tested. GeoCME achieves the best performance among all tested networks.

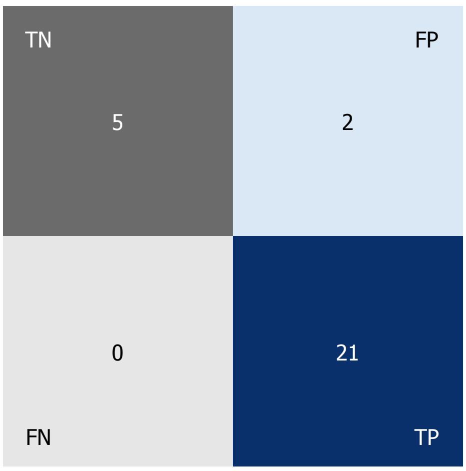

Figure 10. The confusion matrix obtained by GeoCME used as a deterministic prediction model on the test set.

ages, we conducted additional experiments in which we considered the following seven cases.

- E has only LASCO C2 images (denoted C2).
- E has only EIT images (denoted EIT).
- E has only MDI magnetogram images (denoted MDI).
- E has only LASCO C2 and EIT images (denoted C2+EIT).
- E has only LASCO C2 and MDI images (denoted C2+MDI).
- E has only EIT and MDI images (denoted EIT+MDI).
- E has all the images of the three instruments (denoted C2+EIT+MDI).

For each case, we custom-built GeoCME to use the provided data. Figure 11 presents the MCC and TSS results for the seven cases using GeoCME as a deterministic prediction model. Note that the C2+EIT+MDI case in Figure 11 is equivalent to GeoCME in Figure 9. It can be seen in Figure 11 that the combination of LASCO C2, EIT, and MDI images produces the most accurate results with a MCC of 0.807 and a TSS of 0.714 as also shown in Figure 9, indicating that the use of the three types of images together leads to the best performance. When the three types of data are used individually and separately, EIT produces the best results, with a MCC of 0.657 and a TSS of 0.50, followed by MDI, and LASCO C2 is the least effective.

4.3. Probabilistic Forecasting

Our proposed GeoCME can be easily converted from a deterministic prediction model to a probabilistic forecasting model as follows. Instead of comparing the probability (ranging from 0 to 1) produced by the GeoCME model with a pre-determined threshold (which is set to 0.6 in our work), the model simply outputs the probability. For a given CME event, this output now represents a probabilistic estimate of how likely the event will be geoeffective, that is, how likely it will cause a geomagnetic storm with the minimum Dst value less than -50 nT.

We use the Brier score (BS; Brier, 1950) and the Brier skill score (BSS; Wilks, 2010) to assess the performance of a model. The Brier score quantifies the accuracy of the probabilistic forecasts produced by the model by calculating the squared difference between the predicted probabilities and the actual outcomes. Mathematically, the Brier score is calculated by the following formula:

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{4}$$

where N is the number of CME events in the test set, y_i is the actual outcome for the ith event (with 1 representing “geoeffective” and 0 representing “non-geoeffective”), and $\hat{y}_i$ is the predicted probability for the ith event. BS values range from 0 to 1, with a perfect score of 0.

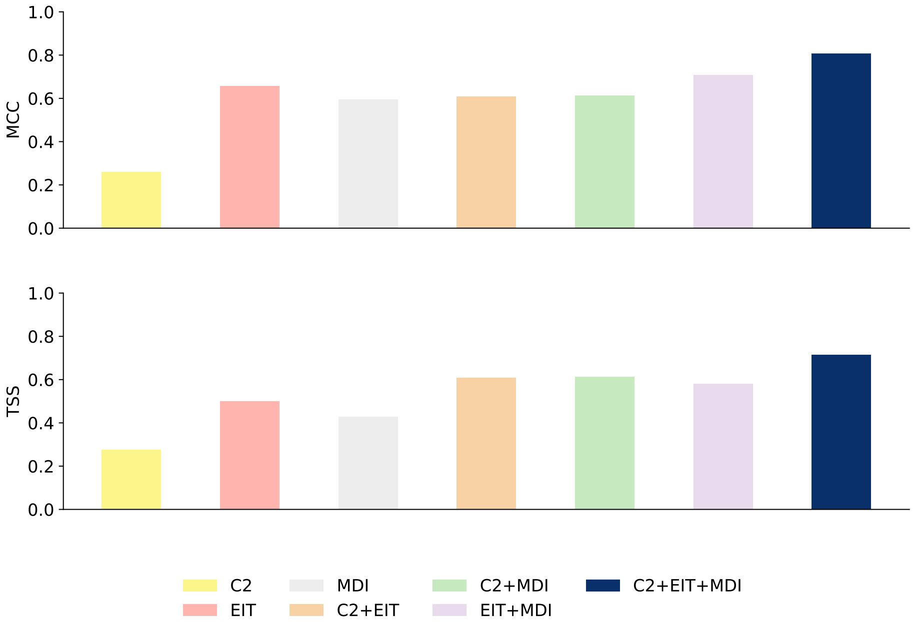

Figure 11. Results of the ablation tests for assessing seven cases (C2, EIT, MDI, C2+EIT, C2+MDI, EIT+MDI, C2+EIT+MDI) using GeoCME as a deterministic prediction model where C2 represents the LASCO C2 images, EIT represents the EIT images, MDI represents the MDI magnetogram images, C2+EIT represents the combination of LASCO C2 and EIT images, C2+MDI represents the combination of LASCO C2 and MDI images, EIT+MDI represents the combination of EIT and MDI images, and C2+EIT+MDI represents the combination of LASCO C2, EIT and MDI images. (Top) MCC of the seven cases tested. (Bottom) TSS of the seven cases tested. C2+EIT+MDI achieves the best performance among all tested cases.

The Brier skill score provides a measure of the model's skill relative to a baseline prediction, calculated as:

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{5}$$

where $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ represents the average of the actual outcomes for the events in the test set. BSS values range from minus infinity to 1, with the perfect score being 1. A BSS of 0 indicates that the model has the same accuracy as the baseline model and a negative BSS indicates that the model performs worse than the baseline.

Figure 12 compares the four networks, namely GeoCME-RN-IRN, GeoCME-RN, GeoCME-IRN, and GeoCME, described in Section 4.2 where the four networks are now used as probabilistic forecasting models. It can be seen in Figure 12 that the GeoCME model again performs the best, achieving the lowest BS of 0.094 and the highest BSS of 0.493. GeoCME-RN-IRN, in which both ResNet and InceptionResNet were removed, performs the worst, as reflected by the highest BS of 0.239 and the lowest BSS of 0.225. These results are consistent

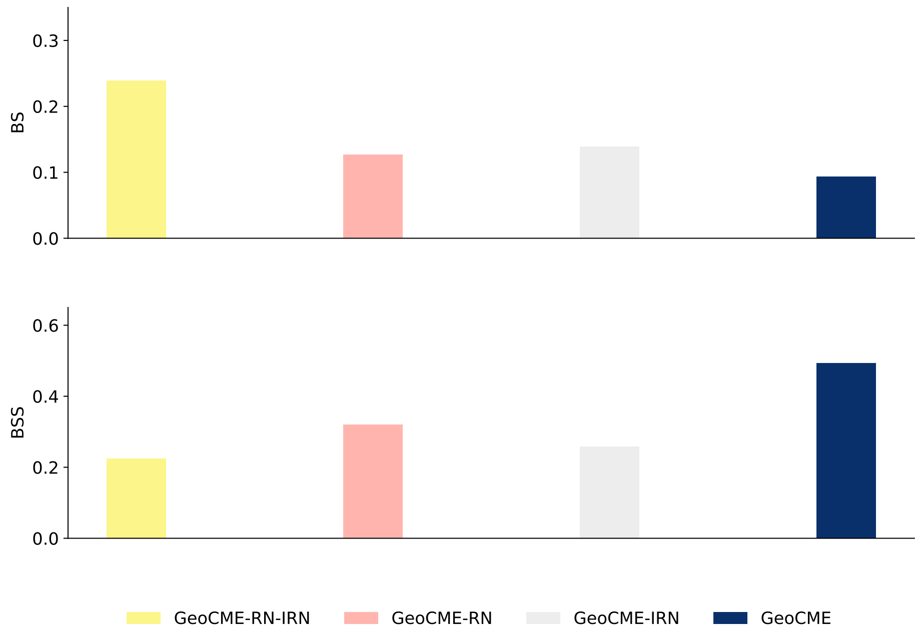

Figure 12. Results of the ablation tests for assessing four networks (GeoCME-RN-IRN, GeoCME-RN, GeoCME-IRN, and GeoCME) used as probabilistic forecasting models where GeoCME-RN-IRN represents GeoCME with RN and IRN models removed, GeoCME-RN represents GeoCME with RN models removed, GeoCME-IRN represents GeoCME with IRN models removed, and GeoCME represents the full model. (Top) BS of the networks tested. (Bottom) BSS of the networks tested. GeoCME achieves the best performance among all tested networks.

with those shown in Figure 9 where the four networks were used as deterministic models.

Figure 13 presents the BS and BSS results for the seven cases (C2, EIT, MDI, C2+EIT, C2+MDI, EIT+MDI, C2+EIT+MDI) defined in Section 4.2, this time using GeoCME as a probabilistic forecasting model. It can be seen in Figure 13 that the combination of LASCO C2, EIT, and MDI images again produces the most accurate results with a BS of 0.094 and a BSS of 0.493, indicating that the use of all data from the three instruments together achieves the best performance. When the three types of data are used individually and separately, EIT yields the best results, with a BS of 0.125 and a BSS of 0.310, followed by MDI, and LASCO C2 is the least effective. These findings are consistent with those shown in Figure 11 where GeoCME was used as a deterministic prediction model.

5. Discussion and Conclusion

We presented GeoCME, a deterministic model that employs ensemble and transfer learning techniques to predict whether a CME event reaching Earth will be geoeffective. Here, a geomagnetic storm is defined as a disturbance of the Earth's

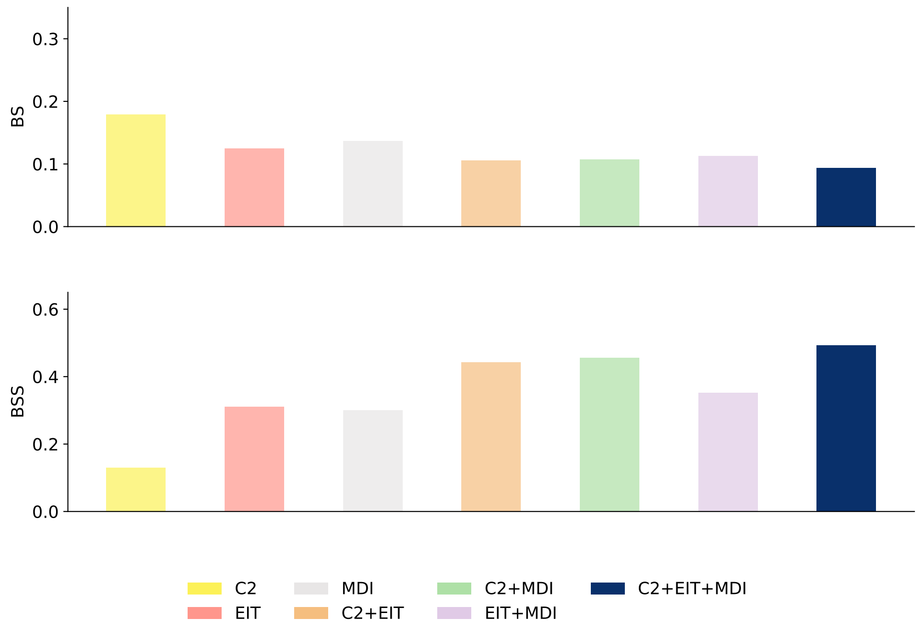

Figure 13. Results of the ablation tests for assessing seven cases (C2, EIT, MDI, C2+EIT, C2+MDI, EIT+MDI, C2+EIT+MDI) using GeoCME as a probabilistic forecasting model where C2 represents the LASCO C2 images, EIT represents the EIT images, MDI represents the MDI magnetogram images, C2+EIT represents the combination of LASCO C2 and EIT images, C2+MDI represents the combination of LASCO C2 and MDI images, EIT+MDI represents the combination of EIT and MDI images, and C2+EIT+MDI represents the combination of LASCO C2, EIT and MDI images. (Top) BS of the seven cases tested. (Bottom) BSS of the seven cases tested. C2+EIT+MDI achieves the best performance among all tested cases.

magnetosphere during which the minimum value of the Dst index is less than -50 nT. Moreover, we converted the deterministic model to a probabilistic forecasting model, which estimates the probability that a CME event will be geoeffective. The GeoCME framework used LASCO C2, EIT 195 Å, and MDI magnetogram images collected by SOHO to make predictions. Our experiments showed that the GeoCME framework can capture the hidden relationships between the SOHO observations and the CME geoeffectiveness, achieving reasonably good performance. Specifically, when used as a deterministic prediction model, GeoCME achieves a MCC of 0.807 and a TSS of 0.714. Approximately 82% of the events in the test set are predicted to be geoeffective. Approximately 8.7% of the predictions are false alarms. When used as a probabilistic forecasting model, GeoCME achieves a BS of 0.094 and a BSS of 0.493. Our experiments also showed that using all three types of solar image together (LASCO C2, EIT, and MDI) performs better than using one or two types of solar image.

We adopted an 80:20 scheme in our dataset that covers CME events from 1997 to 2006 for model training and testing, as described in Section 4.1. In additional experiments, we conducted a five-fold cross-validation to further evaluate the GeoCME framework. Specifically, we divide the dataset into five equally sized subsets or folds, where every two folds have roughly the same number of geoeffective (non-geoeffective, respectively) CMEs. There are 101 geoeffective CMEs and 35 non-geoeffective CMEs in the dataset. Thus, each fold contains approximately 20 geoeffective CMEs and 7 non-geoeffective CMEs. In each run, one fold is used as the test set and the union of the other four folds is used as the training set. There are five folds, and hence five runs. We calculate the average metric values for the five runs. The five-fold cross-validation process yields an average MCC of 0.782 and an average TSS of 0.673 when GeoCME is used as a deterministic prediction model, and an average BS of 0.107 and an average BSS of 0.461 when GeoCME is used as a probabilistic prediction model. Furthermore, in terms of the average metric values, GeoCME outperforms its subnets (GeoCME-RN-IRN, GeoCME-RN, and GeoCME-IRN) and performs the best when all three types of solar image together (LASCO C2, EIT, and MDI) are used in model training and testing. These results are consistent with those obtained from the 80:20 scheme.

Our work relies on existing methods (e.g., Sudar, Vršnak, and Dumbović, 2016; Liu et al., 2018; Amerstorfer et al., 2021; Dumbović et al., 2021; Kaportseva and Shugay, 2021; Baratashvili et al., 2022; Guastavino et al., 2023; Chierichini et al., 2024) to predict whether a CME event would arrive at Earth. When a CME event is predicted to arrive at Earth, we then use the proposed GeoCME to predict whether the CME event will be geoeffective, that is, whether it will cause a geomagnetic storm. Unlike other studies (Besliu-Ionescu et al., 2019; Pricopi et al., 2022), which used CME or solar onset parameters, GeoCME uses solar images to make predictions. The input of GeoCME is composed of directly observed images, which avoids the sophisticated calculation of parameters. Thus, GeoCME has the potential for operational utilization. On the basis of our experimental results, we conclude that GeoCME is a feasible tool for predicting geoeffective CMEs, deterministically or probabilistically.

Acknowledgments The authors thank the members of the Institute for Space Weather Sciences for fruitful discussions. The CME catalog used in this work was created and maintained at the CDAW Data Center by NASA and the Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is an international cooperation project between ESA and NASA.

Author Contribution J.W. and H.W. conceived the study. K.A. wrote the manuscript. All the authors reviewed the manuscript.

Funding K.A. is supported by King Saud University, Saudi Arabia. J.W. and H.W. acknowledge support from NSF grants AGS-2149748, AGS-2228996, OAC-2320147, and NASA grants 80NSSC24K0843 and 80NSSC24M0174. J.J. acknowledges support from NSF grants AGS-2149748 and AGS-2300341. V.Y. acknowledges support from NSF grants AST-2108235, AGS-2114201, AGS-2300341, and AGS-2309939.

Data Availability CME events used in this study were compiled from the RC list in https://izw1.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm and the SOHO/LASCO CME catalog in https://cdaw.gsfc.nasa.gov/CME_list/. SOHO images were collected from the SOHO science archive at https://soho.nascom.nasa.gov/data/archive/.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Abduallah, Y., Jordanova, V.K., Liu, H., Li, Q., Wang, J.T.L., Wang, H.: 2022, Predicting solar energetic particles using SDO/HMI vector magnetic data products and a bidirectional LSTM network. *Astrophys. J. Supp.* **260**, 16. DOI.

Abunin, A.A., Abunina, M.A., Belov, A.V., Chertok, I.M.: 2020, Peculiar solar sources and geospace disturbances on 20-26 August 2018. *Sol. Phys.* **295**, 7. DOI.

Alobaid, K.A., Abduallah, Y., Wang, J.T.L., Wang, H., Jiang, H., Xu, Y., Yurchyshyn, V., Zhang, H., Cavus, H., Jing, J.: 2022, Predicting CME arrival time through data integration and ensemble learning. *Front. Astron. Space Sci.* **9**, 1013345. DOI.

Alobaid, K.A., Abduallah, Y., Wang, J.T.L., Wang, H., Fan, S., Li, J., Cavus, H., Yurchyshyn, V.: 2023, Estimating coronal mass ejection mass and kinetic energy by fusion of multiple deep-learning models. *Astrophys. J. Lett.* **958**, L34. DOI.

Amerstorfer, T., Hinterreiter, J., Reiss, M.A., Möstl, C., Davies, J.A., Bailey, R.L., Weiss, A.J., Dumbović, M., Bauer, M., Amerstorfer, U.V., Harrison, R.A.: 2021, Evaluation of CME arrival prediction using ensemble modeling based on heliospheric imaging observations. *Space Weather* **19**, e02553. DOI.

Augusto, C.R.A., Navia, C.E., de Oliveira, M.N., Nepomuceno, A.A., Raulin, J.P., Tueros, E., de Mendonça, R.R.S., Fauth, A.C., Vieira de Souza, H., Kopenkin, V., Sinzi, T.: 2018, The 2015 summer solstice storm: One of the major geomagnetic storms of solar cycle 24 observed at ground level. *Sol. Phys.* **293**, 84. DOI.

Baker, D.N., Li, X., Pulkkinen, A., Ngwira, C.M., Mays, M.L., Galvin, A.B., Simunac, K.D.C.: 2013, A major solar eruptive event in July 2012: Defining extreme space weather scenarios. *Space Weather* **11**, 585. DOI.

Baratashvili, T., Verbeke, C., Wijsen, N., Poedts, S.: 2022, Improving CME evolution and arrival predictions with AMR and grid stretching in Icarus. *Astron. Astrophys.* **667**, A133. DOI.

Besliu-Ionescu, D., Maris Muntean, G., Dobrica, V.: 2022, Complex catalogue of high speed streams associated with geomagnetic storms during solar cycle 24. *Sol. Phys.* **297**, 65. DOI.

Besliu-Ionescu, D., Talpeanu, D.-C., Mierla, M., Maris Muntean, G.: 2019, On the prediction of geoeffectiveness of CMEs during the ascending phase of SC24 using a logistic regression method. *J. Atmos. Solar-Terr. Phys.* **193**, 105036. DOI.

Brier, G.W.: 1950, Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**, 1. DOI.

Brueckner, G.E., Howard, R.A., Koomen, M.J., Korendyke, C.M., Michels, D.J., Moses, J.D., Socker, D.G., Dere, K.P., Lamy, P.L., Llebaria, A., Bout, M.V., Schwenn, R., Simnett, G.M., Bedford, D.K., Eyles, C.J.: 1995, The Large Angle Spectroscopic Coronagraph (LASCO). *Sol. Phys.* **162**, 357. DOI.

Brueckner, G.E., Delaboudiniere, J.-P., Howard, R.A., Paswaters, S.E., St. Cyr, O.C., Schwenn, R., Lamy, P., Simnett, G.M., Thompson, B., Wang, D.: 1998, Geomagnetic storms caused by coronal mass ejections (CMEs): March 1996 through June 1997. *Geophys. Res. Lett.* **25**, 3019. DOI.

Chertok, I.M.: 2020, On the relationship between the transit time of ICMEs and strength of the initiated geomagnetic storms. *Sol. Phys.* **295**, 74. DOI.

Chierichini, S., Liu, J., Korsós, M.B., Del Moro, D., Erdélyi, R.: 2024, CME arrival modeling with machine learning. *Astrophys. J.* **963**, 121. DOI.

Chollet, F.: 2017, Xception: Deep learning with depthwise separable convolutions. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 1800. DOI.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L.: 2009, ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248. DOI.

Dumbović, M., Čalogović, J., Martinić, K., Vršnak, B., Sudar, D., Temmer, M., Veronig, A.: 2021, Drag-based model (DBM) tools for forecast of coronal mass ejection arrival time and speed. *Front. Astron. Space Sci.* **8**, 58. DOI.

Fry, C.D., Sun, W., Deehr, C.S., Dryer, M., Smith, Z., Akasofu, S.-I., Tokumaru, M., Kojima, M.: 2001, Improvements to the HAF solar wind model for space weather predictions. *J. Geophys. Res.* **106**, 20985. DOI.

Fu, H., Zheng, Y., Ye, Y., Feng, X., Liu, C., Ma, H.: 2021, Joint geoeffectiveness and arrival time prediction of CMEs by a unified deep learning framework. *Remote Sens.* **13**, 1738. 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.* **99**, 5771. DOI.

Goodfellow, I., Bengio, Y., Courville, A.: 2016, *Deep Learning*, MIT Press. http://www.deeplearningbook.org.

Gopalswamy, N.: 2009, Halo coronal mass ejections and geomagnetic storms. *Earth, Planets and Space* **61**, 595. DOI.

Gopalswamy, N., Yashiro, S., Akiyama, S.: 2007, Geoeffectiveness of halo coronal mass ejections. *J. Geophys. Res. (Space Physics)* **112**, A06112. DOI.

Gopalswamy, N., Lara, A., Manoharan, P.K., Howard, R.A.: 2005, An empirical model to predict the 1-AU arrival of interplanetary shocks. *Adv. Space Res.* **36**, 2289. DOI.

Guastavino, S., Candiani, V., Bemporad, A., Marchetti, F., Benvenuto, F., Massone, A.M., Mancuso, S., Susino, R., Telloni, D., Fineschi, S., Piana, M.: 2023, Physics-driven machine learning for the prediction of coronal mass ejections' travel times. *Astrophys. J.* **954**, 151. DOI.

Haines, C., Owens, M.J., Barnard, L., Lockwood, M., Ruffenach, A.: 2019, The variation of geomagnetic storm duration with intensity. *Sol. Phys.* **294**, 154. DOI.

Hayakawa, H., Ebihara, Y., Pevtsov, A.A.: 2024, Analyses of equatorward auroral extensions during the extreme geomagnetic storm on 15 July 1959. *Mon. Not. R. Astron. Soc.* **527**, 7298. 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*, 770. DOI.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: 2017, Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2261. DOI.

Joshi, B., Ibrahim, M.S., Shanmugaraju, A., Chakrabarty, D.: 2018, A major geoeffective CME from NOAA 12371: Initiation, CME-CME interactions, and interplanetary consequences. *Sol. Phys.* **293**, 107. DOI.

Kaportseva, K.B., Shugay, Y.S.: 2021, Use of the DBM model to the predict of arrival of coronal mass ejections to the Earth. *Cosmic Res.* **59**, 268. DOI.

Liu, H., Liu, C., Wang, J.T.L., Wang, H.: 2019, Predicting solar flares using a long short-term memory network. *Astrophys. J.* **877**, 121. DOI.

Liu, H., Liu, C., Wang, J.T.L., Wang, H.: 2020, Predicting coronal mass ejections using SDO/HMI vector magnetic data products and recurrent neural networks. *Astrophys. J.* **890**, 12. DOI.

Liu, J., Ye, Y., Shen, C., Wang, Y., Erdélyi, R.: 2018, A new tool for CME arrival time prediction using machine learning algorithms: CAT-PUMA. *Astrophys. J.* **855**, 109. DOI.

Manoharan, P.K., Gopalswamy, N., Yashiro, S., Lara, A., Michalek, G., Howard, R.A.: 2004, Influence of coronal mass ejection interaction on propagation of interplanetary shocks. *J. Geophys. Res. (Space Physics)* **109**, A06109. DOI.

Martinić, K., Dumbović, M., Čalogović, J., Vršnak, B., Al-Haddad, N., Temmer, M.: 2023, Effects of coronal mass ejection orientation on its propagation in the heliosphere. *Astron. Astrophys.* **679**, A97. DOI.

Mayaud, P.N.: 1980, 2. *What is a geomagnetic index?*, American Geophysical Union (AGU), 2. ISBN 9781118663837. DOI.

Melkumyan, A.A., Belov, A.V., Shlyk, N.S., Abunina, M.A., Abunin, A.A., Oleneva, V.A., Yanke, V.G.: 2024, Forbush decreases and associated geomagnetic storms: Statistical comparison in solar cycles 23 and 24. *Sol. Phys.* **299**, 40. DOI.

Michalek, G., Gopalswamy, N., Lara, A., Yashiro, S.: 2006, Properties and geoeffectiveness of halo coronal mass ejections. *Space Weather* **4**, S10003. DOI.

Mishra, W., Dave, K., Srivastava, N., Teriaca, L.: 2021, Multipoint remote and in situ observations of interplanetary coronal mass ejection structures during 2011 and associated geomagnetic storms. *Mon. Not. R. Astron. Soc.* **506**, 1186. DOI.

Moon, Y.-J., Dryer, M., Smith, Z., Park, Y.D., Cho, K.S.: 2002, A revised shock time of arrival (STOA) model for interplanetary shock propagation: STOA-2. *Geophys. Res. Lett.* **29**, 1390. DOI.

Newell, P.T., Sotirelis, T., Liou, K., Meng, C.-I., Rich, F.J.: 2007, A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *J. Geophys. Res. (Space Physics)* **112**, A01206. DOI.

Pal, S., Nandy, D., Kilpua, E.K.J.: 2022, Magnetic cloud prediction model for forecasting space weather relevant properties of Earth-directed coronal mass ejections. *Astron. Astrophys.* **665**, A110. DOI.

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. DOI.

Pricopi, A.-C., Paraschiv, A.R., Besliu-Ionescu, D., Marginean, A.-N.: 2022, Predicting the geoeffectiveness of CMEs using machine learning. *Astrophys. J.* **934**, 176. DOI.

Raghav, A., Shaikh, Z., Vemareddy, P., Bhaskar, A., Dhamane, O., Ghag, K., Tari, P., Dayanandan, B., Mohammed Al Suti, B.: 2023, The possible cause of most intense geomagnetic superstorm of the 21st century on 20 November 2003. *Sol. Phys.* **298**, 64. DOI.

Richardson, I.G., Cane, H.V.: 2010, Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996 - 2009): Catalog and summary of properties. *Sol. Phys.* **264**, 189. DOI.

Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., Chen, L.: 2018, MobileNetV2: Inverted residuals and linear bottlenecks. In: *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 4510. DOI.

Schrijver, C.J., Kauristie, K., Aylward, A.D., Denardini, C.M., Gibson, S.E., Glover, A., Gopalswamy, N., Grande, M., Hapgood, M., Heynderickx, D., Jakowski, N., Kalegaev, V.V., Lapenta, G., Linker, J.A., Liu, S., Mandrini, C.H., Mann, I.R., Nagatsuma, T., Nandy, D., Obara, T., Paul O'Brien, T., Onsager, T., Opgenoorth, H.J., Terkildsen, M., Valladares, C.E., Vilmer, N.: 2015, Understanding space weather to shield society: A global road map for 2015-2025 commissioned by COSPAR and ILWS. *Adv. Space Res.* **55**, 2745. DOI.

Simonyan, K., Zisserman, A.: 2015, Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations*. http://arxiv.org/abs/1409.1556.

Singh, T., Benson, B., Raza, S.A.Z., Kim, T.K., Pogorelov, N.V., Smith, W.P., Arge, C.N.: 2023, Improving the arrival time estimates of coronal mass ejections by using magnetohydrodynamic ensemble modeling, heliospheric imager data, and machine learning. *Astrophys. J.* **948**, 78. DOI.

Sudar, D., Vršnak, B., Dumbović, M.: 2016, Predicting coronal mass ejections transit times to Earth with neural network. *Mon. Not. R. Astron. Soc.* **456**, 1542. DOI.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: 2016, Rethinking the inception architecture for computer vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2818. DOI.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: 2017, Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: *31st AAAI Conference on Artificial Intelligence*, 4278. https://doi.org/10.1609/aaai.v31i1.11231.

Tan, M., Le, Q.V.: 2019, EfficientNet: Rethinking model scaling for convolutional neural networks. In: *36th International Conference on Machine Learning* **97**, 6105. http://proceedings.mlr.press/v97/tan19a.html.

Telloni, D.: 2022, Statistical methods applied to space weather science. *Front. Astron. Space Sci.* **9**, 865880. DOI.

Vourlidas, A., Patsourakos, S., Savani, N.P.: 2019, Predicting the geoeffective properties of coronal mass ejections: Current status, open issues and path forward. *Philos. Trans. Royal Soc. A* **377**, 20180096. DOI.

Vršnak, B., Gopalswamy, N.: 2002, Influence of the aerodynamic drag on the motion of interplanetary ejecta. *J. Geophys. Res. (Space Physics)* **107**, 1019. DOI.

Wang, Y., Liu, J., Jiang, Y., Erdélyi, R.: 2019, CME arrival time prediction using convolutional neural network. *Astrophys. J.* **881**, 15. DOI.

Wanliss, J.A., Showalter, K.M.: 2006, High-resolution global storm index: Dst versus SYM-H. *J. Geophys. Res. (Space Physics)* **111**, A02202. DOI.

Wilks, D.S.: 2010, Sampling distributions of the Brier score and Brier skill score under serial dependence. *Quarterly Journal of the Royal Meteorological Society* **136**, 2109. DOI.

Wu, C.-C., Liou, K., Lepping, R.P., Hutting, L.: 2019, The 04 - 10 September 2017 Sun-Earth connection events: Solar flares, coronal mass ejections/magnetic clouds, and geomagnetic storms. *Sol. Phys.* **294**, 110. DOI.

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. Supp.* **268**, 69. DOI.

Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O.C., Plunkett, S.P., Rich, N.B., Howard, R.A.: 2004, A catalog of white light coronal mass ejections observed by the SOHO spacecraft. *J. Geophys. Res. (Space Physics)* **109**, A07105. DOI.

Zhang, Z., Shen, C., Chi, Y., Mao, D., Liu, J., Xu, M., Zhong, Z., Wang, C., Wang, Y.: 2023, Comparison of I-ICME and M-ICME fittings and in situ observation parameters for solar cycles 23 and 24 and their influence on geoeffectiveness. *Sol. Phys.* **298**, 138. DOI.

Zhao, X., Dryer, M.: 2014, Current status of CME/shock arrival time prediction. *Space Weather* **12**, 448. DOI.