NRIAG JOURNAL OF ASTRONOMY AND GEOPHYSICS
2025, VOL. 14, NO. 1, 1–9
https://doi.org/10.1080/20909977.2025.2458944

Investigation of double geomagnetic storms on 3 and 4 February 2022 using machine learning approach

Mostafa Hegy[a], Essam Ghamry[a], Ibrahim El-Hamaly[a], Sami Abd El Nabi[b], Ahmad Helaly[b], Adel Fathy[c,d] and T. A. Nahool[e]

[a]Geomagnetic and Geoelectric Department, National Research Institute of Astronomy and Geophysics, Cairo, Egypt; [b]Geophysics Department, Faculty of Science, Ain Shams University, Cairo, Egypt; [c]Physics Department, Faculty of Science, Fayoum University, Fayoum, Egypt; [d]Institute of Basic and Applied Sciences, Egypt-Japan University of Science and Technology (E-JUST), Alexandria, Egypt; [e]Physics Department, Faculty of Science, South Valley University, Qena, Egypt

ABSTRACT

The current work utilising machine learning algorithms to investigate the precursor that follows halo coronal mass ejection (CME), eventually leading to moderate geomagnetic storms on the 3rd and 4th of February 2022. The methodology involved developing and testing a machine learning model on collected data, implemented with a Gradient Boosting Regressor (GBR) technique. The GBR algorithm demonstrated strong performance, yielding high accuracy and precision over various error metrics. These findings underscore the potential of machine learning methods to effectively estimate geomagnetic storms. Specifically, they position the GBR algorithm as an optimal choice for this prediction task, outperforming other evaluated options. This study provides evidence for the suitability of the GBR regressor for reliable SYM-H index modelling. These results show that geomagnetic storms may be directly predicted from solar wind parameter data, with a lead time of several days for forecasting, which is important for improving space weather forecasts. According to the study, using the GBR regressor model improves the performance to up to 95%.

ARTICLE HISTORY
Received 28 October 2024
Revised 3 January 2025
Accepted 19 January 2025

KEYWORDS
Machine learning algorithms; forecasting magnetic field; Gradient Boosting Regressor (GBR) regressor; geomagnetic storms

1. Introduction

The Earth's magnetic field interacts with charged particle currents carried by the solar wind to form geomagnetic storms. The most frequent cause of geomagnetic storms is a region of weak magnetic field on the surface of the Sun called coronal mass ejections (CMEs), which are massive expulsions of plasma and magnetic field from the Sun's corona or a corotating interaction in a zone dominated by the solar wind Gonzalez et al. (1999). Solar disturbances produce a charged particle stream that is carried by the solar wind. The exact speed of the solar wind and the southern direction of the interplanetary magnetic field (IMF) determine the extent of the interaction between the Earth's magnetosphere and anomalies of the solar Varela et al. (2022). It is commonly recognised that the stronger the interaction, the greater the fluctuation in the Bz component of the IMF and solar wind speed. The zonal equatorial electric field component can occasionally be seen to reverse when the IMF shifts from a constant southerly direction to a northward one; the storm is associated with a rapid change in the magnetosphere's convective electric field Chakrabarty et al. (2017). A sequence of solar events, including an M1.0 class flare and accompanying coronal mass ejections (CMEs), set off geomagnetic storms that occurred on 3rd and 4 February 2022. Increased air density and ionospheric disruptions were caused by these storms' substantial effects on Earth's atmosphere and ionosphere Dang et al. (2022). Thirty-eight Starlink satellites were destroyed by storms, underscoring the necessity for improved knowledge and forecasting of space weather. These results are in line with other studies on the impacts of geomagnetic storms on the magnetosphere and ionosphere of Earth Lakhina and Tsurutani (2018).

Space weather prediction is the process of estimating the strength of the onset of a geomagnetic storm using solar and interplanetary data Srivastava (2005). A geomagnetic storm causes significant modifications to the Earth's magnetic field auroras, and electric currents. These changes can be harmful to human life and health and can accelerate charged particles in the interplanetary (IP) space and enhance them Siscoe and Schwenn (2006). Space weather is often characterised by geomagnetic storms. To accomplish this, a prediction technique based on the solar and IP features of geo-effective CMEs must be developed. The primary subject in the context of space weather today

CONTACT Mostafa Hegy mostafa.hegy@nriag.sci.eg Geomagnetic and Geoelectric Department, National Research Institute of Astronomy and Geophysics, El Marsed Street 1, Cairo, Helwan 11421, Egypt

is the potential forecasting of a geomagnetic storm. The truth is that a great deal of research has been done to determine the precise relationship between solar processes, interplanetary events and geomagnetic activity Zhang et al. (2022) and Li et al. (2007). However, several studies concentrated on forecasting the disturbance storm time (Dst) index Temerin and Li (2002), Gonzalez et al. (2004), Pallocchia et al. (2006) which evaluate the solar activity-driven dynamic of the symmetric portion of the ring current, Sugiura and Chapman (1960). Space weather affects electrical power distribution systems, communication networks, satellites and terrestrial radars; hence, accurate, fast and reliable machine learning methods of forecasting are crucial for mitigating the negative effects caused by geomagnetic storms Akasofu et al. (2021).

Machine learning (ML) has witnessed great potential to forecast the geomagnetic storm. Several authors have used solar wind data and input parameters, while others have used solar features such as sunspot number and size of coronal hole to reach a good performance. In the 1980s, attempts to anticipate geomagnetic indices by Mayaud (1980) through a linear prediction model. The SYM-H index indicates the ring current's asymmetric and longitudinally symmetric components, and an attempt is made to use ANN to create a prediction model for this index Bhaskar and Vichare (2019). They used the Autoregressive network (NARX) to implement the SYM-H index for 92 geomagnetic storms between 1998 and 2013. The SYM-H index-trained network predicts SYM-H very well, and the average correlation between observed and predicted SYM-H is strong (R^2 ~0.88), meaning that the network models nearly 77% of the variability in SYM-H. For this reason, we used different models to enhance the correlation coefficient.

In this work, we assess the performance of machine learning approaches in modelling, specifically Gradient Boosting Regressor (GBR) to predict the SYM-H index. Analysis showed the GBR algorithm delivered a good performance, attaining a strong predictive accuracy of 95%, exceeding alternative evaluated methods. By leveraging GBR for SYM-H index forecasting, the high precision demonstrates machine learning, and GBR in particular, as an effective modern approach for modelling this phenomenon. The successful application and predictive capabilities achieved underscore the suitability of machine learning for advancing insights into SMY-H through data-driven modelling. The structure of the paper is organised into the following sections: Section 2 explains the data sets, Section 3 demonstrates the methodology, Section 4 presents the results, Section 5 discusses our findings, and finally the last section is the conclusion.

2. Data sets

2.1. Parameters describing the geomagnetic storms

The solar wind data used in the current study to characterise the geomagnetic disturbance are solar wind velocity (v), proton density (n), proton temperature (T), flow pressure (p), electric field (Ey), magnetic field (Bz) and the symmetric disturbance storm time index (SYM-H). These data have been obtained from the OMNI web data services interface, which is available through this link: (https://omniweb.gsfc.nasa.gov.). The data are modified (Level −3) High-Resolution OMNI (HRO) data, derived from the conclusive Wind plasma data, in 1-min resolution Zhang et al. (2022). The solar wind can be described as a two-component flow with fast, tenuous, quiescent flow emanating from coronal holes Riley et al. (2011). In this study, we used four days of solar parameter and SYM-H index data. These data represent days (1, 2, 3, and 4) in February 2022. In this study, we used four days solar parameter and SYM-H index data. These data represent days (1, 2, 3 and 4) February 2022.

2.2. Ground observatories data

The ground magnetic observatories KAK, BSL, ABK, and LER are situated at the following locations: KAK station is in Kanoya, Japan, at Latitude 36.232 and Longitude 140.186. BSL station is found in Baker Lake, Canada, with coordinates Latitude 64.318 and Longitude 263.988 Turbitt et al. (2003). ABK station is situated in Abisko, Sweden, at Latitude 68.358 and Longitude 18.823 (Mandrikova and Rodomanskay 2021). LER station located in Lerwick, United Kingdom, with coordinates Latitude 60.138 and Longitude 358.817 Turbitt et al. (2003). Data from these observatories have been sourced from the British Geological Survey (BGS) website (https://geomag.bgs.ac.uk/.) Turbitt et al. (2003). The selected data used in our analysis were downloaded from this site and are monitored through these stations.

2.3. Misallat observatory in Egypt data

We used geomagnetic raw data obtained from Misallat observatory (MLT) operated by Geomagnetism Laboratory, the National Research Institute of Astronomy and Geophysics (NRIAG), Egypt El-Eraki et al. (2014); Ghamry et al. (2016). MLT observatory was established in 1960 as a substitute for the geomagnetic observations that had been running since 1903 at the main quartier of NRIAG at Helwan. MLT is

located at 29° 30′ 52″ N and 30° 53′ 22″ E. The observatory at Misallat has a Magson digital fluxgate magnetometer, an overhauser magnetometer for total field measurements, and a single-axis DI fluxgate sensor mounted on a nonmagnetic theodolite for absolute measurements.

3. Methodology

GBR is a powerful machine learning algorithm that combines multiple weak regression models to create a highly accurate predictive model Konstantinov and Utkin (2021). GBR works by iteratively training a sequence of simple regression models, each of which attempts to minimise the residual error from the previous model. This algorithm is highly robust to noisy data and can handle complex interactions between features without the need for extensive data preprocessing or parameter tuning Guelman (2012). GBR has been successfully applied to a variety of regression problems, including predicting surface area, pore volume, and yield in biomass-derived activated carbon production, with R-squared values ranging from 0.76 to 0.91 Zou et al. (2024).

Researchers have also explored extensions of GBR, such as wide boosting, which inserts a matrix multiplication between the output of the GBR model and the loss function, allowing for increased model complexity and improved predictive performance Horrell (2020). However, GBR models can be computationally intensive, especially for large datasets, and may require careful tuning of hyperparameters to achieve optimal performance Zheng et al. (2017). The explainability of GBR predictions, especially with earth observation data, is still an open issue that requires more focus by researchers Okolie et al. (2023). In the domains of regression and classification tasks, gradient boosting surpasses other methods like random forests or support vector machines in both accuracy and efficiency. Its capability to amalgamate multiple weak learners into a potent ensemble of strong learners makes it particularly suitable for large-scale applications Bentéjac et al. (2021). The statistical RMSE, MAE and R^2 values used in the study are calculated as shown in Eqs. 1, 2, 3 and 4, respectively:

$$\boldsymbol{RMSE} = \sqrt{\boldsymbol{MSE}} \tag{1}$$

$$\boldsymbol{MAE} = 1/\boldsymbol{n}\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}\left(\left|\boldsymbol{y_i} - \widehat{\boldsymbol{y}_i}\right|\right) \tag{2}$$

$$\boldsymbol{R}^2 = 1 - (\boldsymbol{SSR}/\boldsymbol{SST}) \tag{3}$$

Where:

- RMSE is the Root Mean Squared Error.
- MAE is the Mean Absolute Error.
- R^2 is the correlation coefficient.
- n is the total number of samples in the dataset.
- $y_i$is the actual (true) value of the target variable for the i-th sample.
- $\widehat{y_i}$ is the predicted value of the target variable for the i-th sample.
- SSR is the sum of squared residuals, which is a measure of the error between the predicted values and the actual values.
- SST is the total sum of squares, which is a measure of the total variation in the data.

The R^2 value can also be calculated using the following formula:

$$\boldsymbol{R}^2 = (\boldsymbol{SST} - \boldsymbol{SSR})/\boldsymbol{SS} \tag{4}$$

In both cases, the R^2 value is a number between 0 and 1, where 1 indicates a perfect fit and 0 indicates no fit. A higher R^2 value indicates a better fit.

4. Results

4.1. Geomagnetic storms analysis

The key indices for geomagnetic storm analysis are the Dst and SYM-H indices, which measure the variation of the north-south component of the Earth's magnetic field Katus and Liemohn (2013). We analysed the SYM-H index for double geomagnetic storms on 3 and 4 February 2022. The locations of ground magnetic observatories are shown in (Figure 1), and we used (KAK, MLT, BSL, ABK, LER) stations to analyse these events. Figure 2 illustrates the variation of the solar parameters of the double storm on 3rd and 4 February 2022. Figure 2(a) shows the solar wind dynamic pressure; it was an increase in the dynamic pressure of about 18 and 14 nPa on 3rd and 4th February, respectively. Figure 2(b) explains the solar wind velocity. Also, the velocity increased to around − 420 km/s on day 3 and − 450 km/s on day 4 in February. We observed a rise in solar wind density of about (35, 20 cm-3), which was observed on 3rd and 4th of February, respectively, in (Figure 2(c)). The interplanetary magnetic field magnitude (IMF) increased, ranging from 13 to 18 (nT) on 4th and 3rd February, respectively (Figure 2(d)). The range of change in magnetic activity from 62 to 66 nT is shown in SYM-H in (Figure 2(e)).

Ground magnetic observatories during geomagnetic storms are shown in (Figure 3). Figure 3a shows the (LER) station located at 60.1380N, 1.1830W for the British Geological Survey (BGS), United Kingdom. Figure 3(b) is the MLT station in Egypt, and it is located at 29°30'51"N 30°53'22"E. The (BSL) station is located at 30.3500 N, 89.6400 W for the United States Geological Survey (USGS), United States of America in (Figure 3c). The fourth station is the (ABK); it is located at 68.3580 N, 18.8230 E for the Geological Survey of Sweden (SGU), Sweden. The last

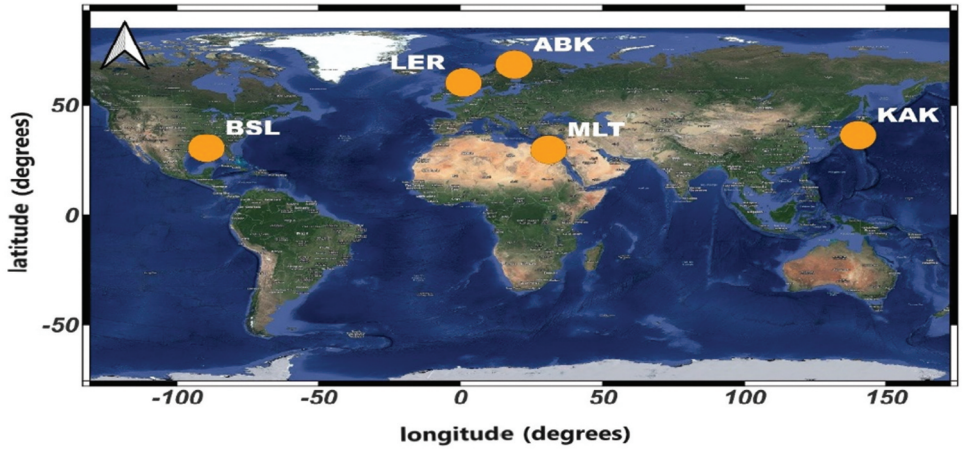

Figure 1. Locations of ground magnetic observatories.

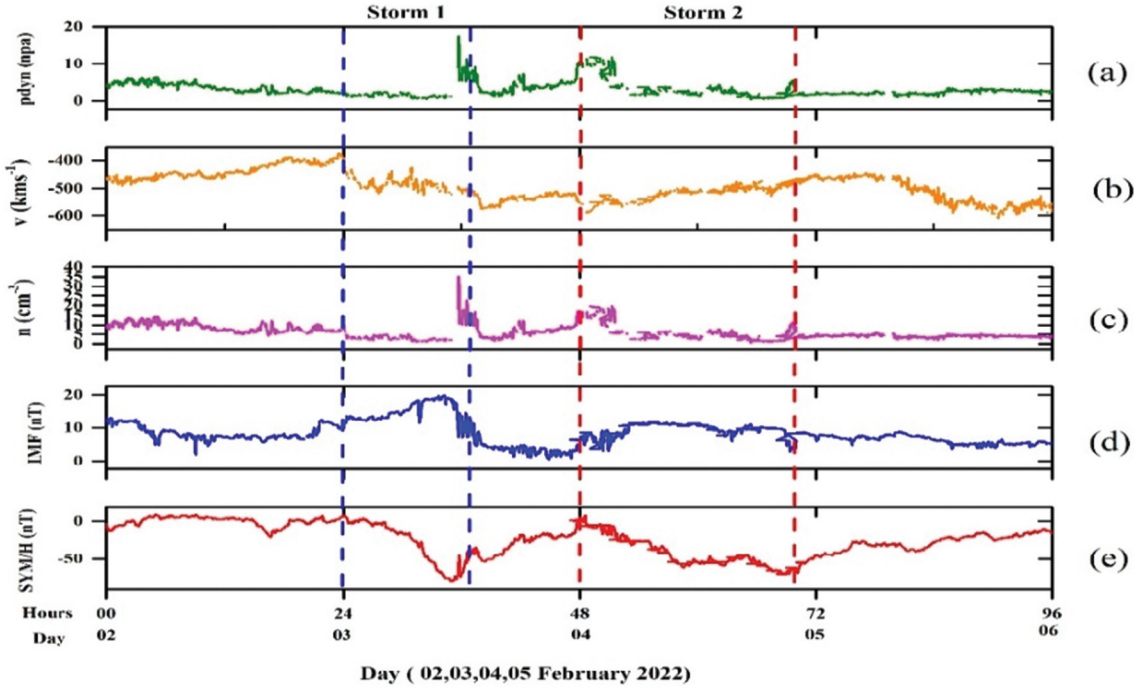

Figure 2. Solar wind parameters during the geomagnetic storms on 3rd and 4th February 2022.

station is (KAK), located at 36.2320 N, 140.1860 E and belong to the Japan Meteorological Agency (JMA), Japan.

4.2. Machine learning analysis using the GBR algorithm

The input variables for the GBR model should include solar wind parameters such as solar (v), (n), (T), (p), (Ey), (Bz) as these have been identified as having the greatest impact on the output variable, the geomagnetic index (SYM-H). Table 1 presents a descriptive statistical overview of the key input and output variables in the SYM-H index data used to train and test the GBR model. We used the machine learning model that was developed and evaluated using a total of 4369 data points in Table 1. Table 2 presents the performance metrics of the GBR model on both the training data and testing data for predicting the SYM-H index. The GBR model achieves low MAE and RMSE scores of 3.79 and 5.20, respectively, indicating its strong generalizability. The model is evidenced by the high R^2 values of 0.957 for testing data and 0.950 for training data, demonstrating excellent predictive performance.

Figure 4 shows the residual values calculated from the actual observed SYM-H index in the blue column, the predicted SYM-H index values in the red column, and the error values in the green column. We used the GBR model to estimate the residuals of the actual and predicted SYM-H index. The residuals of the actual and predicted SYM-H index using the GBR model,

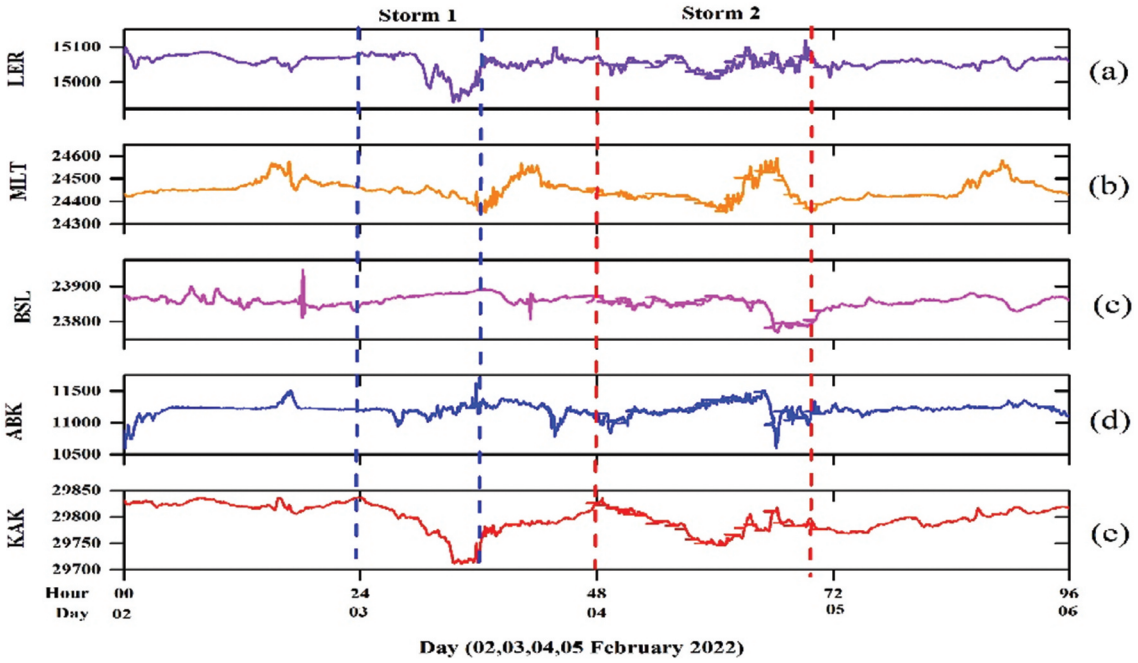

Figure 3. Ground magnetic observatories during the geomagnetic storms on 3rd and 4th February 2022.

Table 1. Descriptive statistics and data analysis of solar wind parameters and SMY-H index.

	IMF	V_X	N	npa	SMY-H
Count	4369.00000	4369.000000	4369.000000	4369.000000	4369.000000
mean	7.97412	492.083566	6.292889	3.036576	23.427329
std	3.44962	49.077159	3.656215	1.888788	21.765353
min	0.49000	608.600000	0.810000	0.400000	79.000000
25%	5.64000	530.700000	3.790000	1.860000	40.000000
50%	7.46000	490.000000	4.910000	2.540000	20.000000
75%	9.79000	457.500000	8.220000	3.710000	3.000000
Max	19.85000	373.200000	35.120000	17.500000	9.000000

Table 2. Performance of our machine learning model (GBR) for training and testing data for predicting SYM-H index.

Metric	Training Data	Testing Data
MAE	3.79	3.84
RMSE	5.20	5.03
R^2	0.950	0.957

about 2.77, are calculated through the following equation:

$$\textit{Residual} = \textit{Observed} - \textit{Predicted} \tag{5}$$

The "Error" column quantifies the difference between the actual and predicted SYM-H index values. Figure 5a displays the variables Bz, V, n, T, P, Ey and SYM-H, which are indicated on both the horizontal and vertical axes. The correlation coefficient between the variables in each row and column is represented by a cell in the matrix. Figure 5b illustrates the significance of every solar wind parameter and SMY-H index to the GBR model. The data set many characteristics shown on the x-axis. Each parameter relevance score is shown on the y-axis, where higher scores denote more significance for the

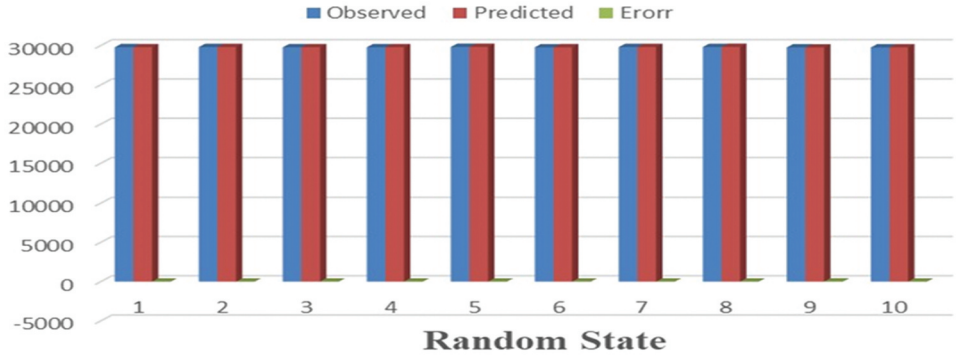

Figure 4. Residuals of actual and predicted SMY-H index using the GBR model.

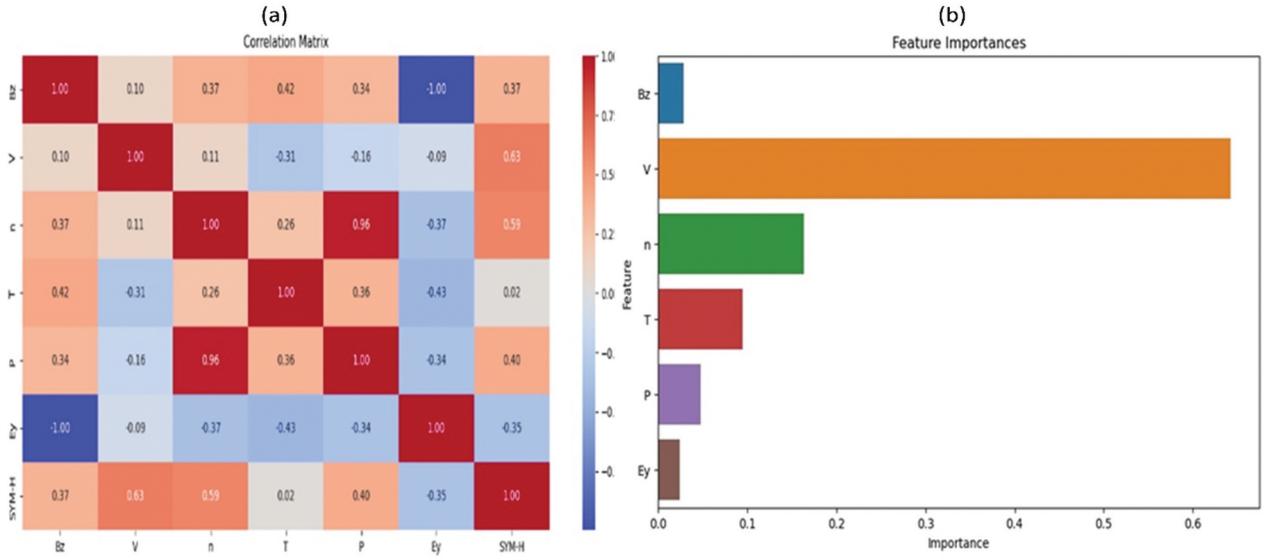

Figure 5. (a) Correlation heatmap of dataset variables used on predicting SYM-H index, (b) Typical features importances dataset variables used on predicting SYM-H index.

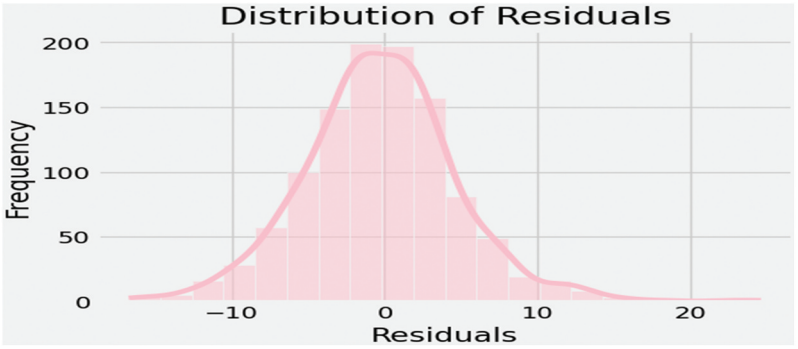

Figure 6. Distribution of error values between observed and predicted values of SYM-H index using GBR algorithm.

correctness of the model. The range of scores is from 0 to 1. Velocity (v) appears to be the most crucial component, with a correlation of 0.63; so the velocity has the most impact on the target variable due to its dominating score.

Figure 6 displays the low values of the GBR model's error. The histogram shows the frequency on the y-axis and residuals on the x-axis. The smooth curve represents the probability density estimate of the residuals. The range of the distribution of residuals ranged from −16.75 to + 24.60. Observed and predicted values of the SYM-H index for the GBR model are shown in (Figure 7). The blue line represents predicted values plotted against observed values for the SYM-H index. The red line indicates where predicted values exactly equal observed values. The closer

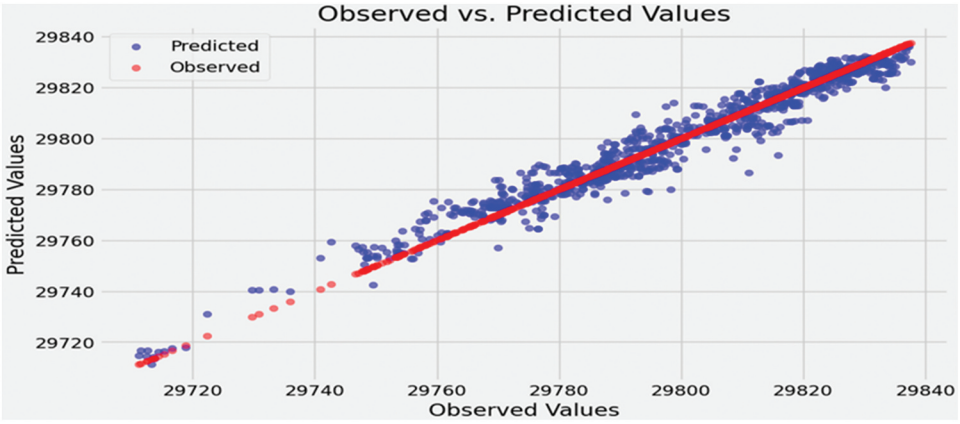

Figure 7. Observed and predicted values of SYM-H index using the GBR algorithm.

Figure 8. Training and test loss over alterations for GBR model.

the blue points are to this line, the better the GBR model's predictions. The blue points are tightly clustered along the red line, indicating that the predicted values closely match the observed values.

Figure 8 shows the training loss of the model exhibits a consistent lower trend with increasing training, suggesting that the model is getting better at fitting the training data. The loss of the GBR model against iterations for training and against iterations for the testing data are shown left and right, respectively, in (Figure 8). The GBR training loss shows a steady downward trend as training progresses and an alteration of about 180, indicating that the model is continuously improving to fit the training data. Simultaneously, the test loss also decreases over time, demonstrating the model's ability to generalise its learning to new, unfamiliar data. The simultaneous reductions in training and test loss highlight the model's capacity to learn underlying patterns from the training data that transfer well to making accurate predictions on new data. The predicted values closely match the actual values, demonstrating the model's effectiveness in accurately forecasting the magnetic field for both sets of training and testing datasets in (Figure 9).

5. Discussion

GBR has emerged as a powerful tool for predicting geomagnetic storms using machine learning techniques to increase prediction accuracy. The GBR model is trained using a combination of solar wind patterns, interplanetary magnetic field (IMF) data, and SYM-H index. In this paper, we used the GBR algorithm to predict the SYM-H index. This model demonstrates excellent predictive performance. Recent research highlights how the GBR was successfully used to predict the SYM-H index, which is important for assessing geomagnetic storm activity. The GBR model is trained using a combination of solar wind patterns, interplanetary magnetic field (IMF) data, and SYM-H index. This approach allows for capturing the intricate dynamics that lead to geomagnetic storms Iong et al. (2022), Sierra-Porta et al. (2024). Studies have shown that GBR outperforms traditional forecasting methods and some neural network methods in predicting accuracy. For example, Iong et al. (2022) showed that the GBR provided a statistically significant improvement in root square error compared to the existing model. GBR has shown superior

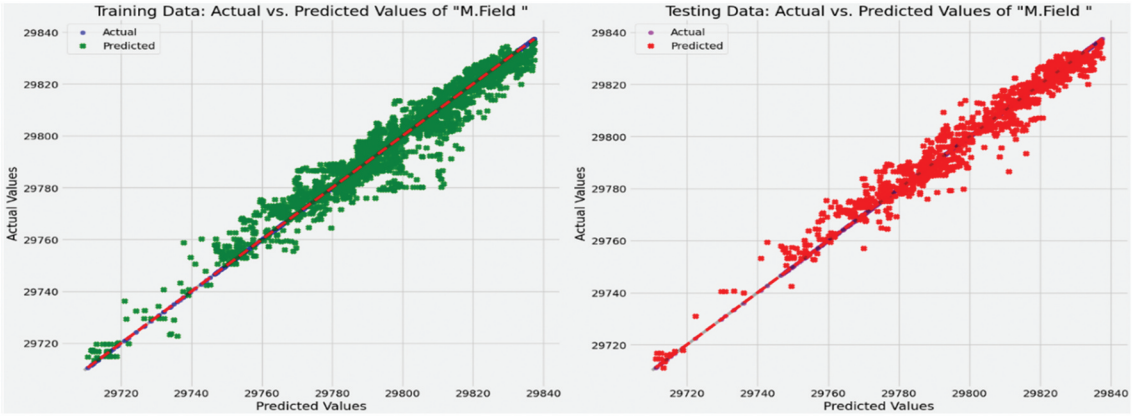

Figure 9. Actual and predicted values for training and test data of SYM-H index using (GBR) algorithm.

predictive accuracy compared to other machine learning models such as random forests and traditional statistical methods. It accurately identifies important geomagnetic activity, such as solar wind speed and magnitude, that are important in storm prediction Balaji and Ranganathan (2024), Sierra-Porta et al. (2024). Through previous studies, in this study we used the GBR model because it has high accuracy for forecasting geomagnetic storms.

Several scientists have been using ML models to forecast geomagnetic storms using data from the solar wind (SW) and geomagnetic indexes. Barkhatova et al. (2008) used artificial neural networks (ANN) to investigate two indexes of SYM-AU and ASY-AL. They showed that the correlation coefficient (>50%) between the real and retrieved sequences is observed. Luo et al. (2018) assessed the performance of GPS under different geomagnetic storm conditions during solar cycle 24 using (ANN). According to statistical findings, the increased correlation coefficients under moderate, powerful, and superstorm circumstances are 13.04%, 56.52%, and 69.57%, respectively. Koklu (2022) estimated geomagnetic indices, including Dst and Kp, using an ANN model. With an accuracy of over 90%, the discussion is reliable in estimating geomagnetic indices. Basciftci (2023) investigated the four moderate geomagnetic storms that occurred in 2015 using ANN to predict (Dst, Ap, and AE). Through their result, the accuracy for ANN is 89%. Budiman et al. (2023) used Random Forest regression to forecast the Kp index, and the efficiency in this model is 85.55%.

Overall, our results matching previous studies using the metrics validate the strong performance of the GBR model by accurately estimating the SYM-H index. Which is important for assessing geomagnetic storm intensity in Table 2, the training data used to build the model, the GBR model achieves low MAE and RMSE scores of 3.79 and 5.20, respectively. This suggests the model fits the training data very well. The testing data provides an out-of-sample evaluation of model performance. The GBR model achieves similar MAE and RMSE values on the testing data, indicating its strong generalisability. The model demonstrates outstanding predictive performance as it explains 95–96% of the variation in magnetic field, as evidenced by the high (R^2) values of 0.957 for testing data and 0.950 for training data, demonstrating excellent predictive performance. GBR offers significant advantages for forecasting geomagnetic storms; it is essential to remain vigilant about its limitations. By implementing strategies to mitigate these challenges and conducting a thorough self-assessment throughout the modelling process, we can enhance the reliability and effectiveness of our predictions in this critical area of research.

6. Conclusion

We utilise machine learning algorithms to investigate the precursor that follows halo coronal mass ejection (CME), eventually leading to moderate geomagnetic storms on the 3rd and 4th of February 2022. The methodology involved developing and testing a machine learning model on collected data, implemented with GBR technique. This study provides evidence for the suitability of the GBR regressor for forecasting the SYM-H index. According to the study, using the GBR regressor model gives performance up to 95 %. This load to the GBR algorithm is one of the best machine learning algorithms, and it gives high accuracy for our problem and showed outstanding capabilities, producing high levels of precision and accuracy across a range of error measures. This work is a preliminary study that should be extended by extra studies that may discuss the width of machine learning to the prediction of geomagnetic storms.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Akasofu SI. 2021. A review of studies of geomagnetic storms and auroral/magnetospheric substorms based on the electric current approach. Front Astron and Space Sci. 7:604750.

Balaji SRA, Ranganathan P. 2024 Jan. Geomagnetic storm forecasting using machine learning models. 2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC); Jan 81–0; Las Vegas, NV, USA. p. 0513–0520. https://ieeexplore.ieee.org/xpl/conhome/10426790/proceeding.

Barkhatova OM, Revunov SE, Barkhatov NA, Levitin AE. 2008. Polar current systems at the main phase of a geomagnetic storm. Proc. 31st Annual Seminar. Apatity; Feb 26–29; Apatity, Russia. p. 13–16.

Basciftci F. 2023. Using artificial neural networks in the investigation of four moderate geomagnetic storms (mGSs) that occurred in 2015. Adv Space Res. 71 (10):4382–4400. doi: 10.1016/j.asr.2023.01.001.

Bentéjac C, Csörgő A, Martínez-Muñoz G. 2021. A comparative analysis of gradient boosting algorithms. Artif Intel Rev. 54(3):1937–1967. doi: 10.1007/s10462-020-09896-5.

Bhaskar A, Vichare G. 2019. Forecasting of SYMH and ASYH indices for geomagnetic storms of solar cycle 24 including st. Patrick's day, 2015 storm using NARX neural network. J Space Weather Space Clim. 9:A12. doi: 10.1051/swsc/2019007.

Budiman B, Alamsyah N, Yoga TP, Alamsyah RYR, Habibi C. 2023. Machine learning-based classification of house prices: a comparative study. 2023 5th International Conference on Cybernetics and Intelligent System (ICORIS); Oct 6–7; Pangkalpinang, Indonesia. IEEE. p. 1–6. https://ieeexplore.ieee.org/xpl/conhome/10352108/proceeding.

Chakrabarty D, Hui D, Rout D, Sekar R, Bhattacharyya A, Reeves GD, Ruohoniemi JM. 2017. Role of IMF by in the prompt electric field disturbances over equatorial ionosphere during a space weather event. J Geophys Res: Space Phys. 122(2):2574–2588. doi: 10.1002/2016JA022781.

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. Space Weather. 20(8):e2022SW003152. doi: 10.1029/2022SW003152.

El-Eraki MA, Rabeh T, Lethy A, Ghamry E, Samy A. 2014. Detection of long period Pc5 pulsation during 01st/Jul/2011 to 30th/Jun/2012 recorded at Egyptian and some INTERMAGNET observatories. NRIAG J Astron Geophys. 3(2):192–197. doi: 10.1016/j.nrjag.2014.11.001.

Ghamry E, Lethy A, Arafa-Hamed T, Abd Elaal E. 2016. A comprehensive analysis of the geomagnetic storms occurred during 18 February and 2 March 2014. NRIAG J Astron Geophys. 5(1):263–268. doi: 10.1016/j.nrjag.2016.03.001.

Gonzalez WD, Dal Lago A, De Gonzalez AC, Vieira LEA, Tsurutani BT. 2004. Prediction of peak-dst from halo CME/magnetic cloud-speed observations. J Atmos Sol-Terr Phys. 66(2):161–165. doi: 10.1016/j.jastp.2003.09.006.

Gonzalez WD, Tsurutani BT, Clúa de Gonzalez AL. 1999. Interplanetary origin of geomagnetic storms. Space Sci Rev. 88(3):529–562. doi: 10.1023/A:1005160129098.

Guelman L. 2012. Gradient boosting trees for auto insurance loss cost modeling and prediction. Expert Syst Appl. 39(3):3659–3667.. doi: 10.1016/j.eswa.2011.09.058.

Horrell MT. 2020. Wide boosting. arXiv preprint arXiv:2007.09855.

Iong D, Chen Y, Toth G, Zou S, Pulkkinen T, Ren J, Camporeale E, Gombosi T. 2022. New findings from explainable SYM-H forecasting using gradient boosting machines. Space Weather. 20(8):e2021SW002928. doi: 10.1029/2021SW002928.

Katus RM, Liemohn MW. 2013. Similarities and differences in low- to middle-latitude geomagnetic indices. J Geophys Res: Space Phys. 118(8):5149–5156. doi: 10.1002/jgra.50501.

Koklu K. 2022. Using artificial neural networks for comparison of the 09 March 2012 intense and 08 May. Adv Space Res. 70(10):2929–2940. doi: 10.1016/j.asr.2022.07.067.

Konstantinov AV, Utkin LV. 2021. Interpretable machine learning with an ensemble of gradient boosting machines. Knowl-Based Syst. 222:106993. doi: 10.1016/j.knosys.2021.106993.

Lakhina GS, Tsurutani BT. 2018. Supergeomagnetic storms: past, present, and future. In: Buzulukova N, editor. Extreme events in geospace. Extreme Events in Geospace: Elsevier. p. 157–185. https://www.sciencedirect.com/science/article/abs/pii/B9780128127001000078?via%3Dihub.

Li X, Oh KS, Temerin M. 2007. Prediction of the AL index using solar wind parameters. J Geophys Res: Space Phys. 112(A6). doi: 10.1029/2006JA011918.

Luo X, Gu S, Lou Y, Xiong C, Chen B, Jin X. 2018. Assessing the performance of GPS precise point positioning under different geomagnetic storm conditions during solar cycle 24. Sensors. 18(6):1784. doi: 10.3390/s18061784.

Mandrikova O, Rodomanskay A. 2021 Sep. Method for detecting geomagnetic disturbances based on the wavelet model of geomagnetic field variations. 2021 International Conference on Information Technology and Nanotechnology (ITNT); Sep 20–24; Samara, Russian Federation, IEEE. p. 1–7. https://ieeexplore.ieee.org/xpl/conhome/9648913/proceeding.

Mayaud PN. 1980. Derivation, meaning, and use of geomagnetic indices Vol. 22, 154. Washington (DC): American Geophysical Union.

Okolie C, Mills J, Adeleke A, Smit J, Maduako I. 2023. The explainability of gradient-boosted decision trees for digital elevation model (DEM) error prediction. Int Archiv Photogrammetry, Remote Sens Spatial Inf Sci. 48:161–168. doi: 10.5194/isprs-archives-XLVIII-M-3-2023-161-2023.

Pallocchia G, Amata E, Consolini G, Marcucci MF, Bertello I. 2006. May. Geomagnetic D_{st} index forecast based on IMF data only. Ann Geophysicae. 24(3):989–999. Göttingen, Germany: Copernicus Publications.10.5194 / angeo-24-989-2006.

Riley P, Lionello R, Linker JA, Mikic Z, Luhmann J, Wijaya JZ. 2011. Global MHD modeling of the solar corona and inner heliosphere for the whole heliosphere interval. Sol Phys. 274(1–2):361–377. doi: 10.1007/s11207-010-9698-x.

Sierra-Porta D, Petro-Ramos JD, Ruiz-Morales DJ, Herrera-Acevedo DD, García-Teheran AF, Alvarado MT. 2024. Machine learning models for predicting geomagnetic storms across five solar cycles using dst index and heliospheric variables. Adv Space Res. 74(8):3483–3495. doi: 10.1016/j.asr.2024.08.031.

Siscoe G, Schwenn R. 2006. CME disturbance forecasting. Space Sci Rev. 123(1–3):453–470. doi: 10.1007/s11214-006-9024-y.

Srivastava N. 2005. Nov. Predicting the occurrence of superstorms. Ann Geophysicae. 23(9):2989–2995. doi: 10.5194/angeo-23-2989-2005.

Sugiura M, Chapman S. 1960. The average morphology of geomagnetic storms with sudden com-mencement, Abh. *Akad. Wiss. Gottingen, math. Phys. Kl. $ onderhe t.* 4.

Temerin M, Li X. 2002. A new model for the prediction of dst on the basis of the solar wind. J Geophys Res: Space Phys. 107(A12):SMP–31. doi: 10.1029/2001JA007532.

Turbitt CW, Flower SM, Riddick JC. 2003. The united kingdom geomagnetic observatory network: disseminating one-second data in near real-time.

Varela J, Brun AS, Strugarek A, Réville V, Zarka P, Pantellini F. 2022. MHD study of the planetary magnetospheric response during extreme solar wind conditions: earth and exoplanet magnetospheres applications. Astron Astrophys. 659:A10. doi: 10.1051/0004-6361/202141181.

Zhang H, Xu HR, Peng G, Qian YD, Zhang XX, Yang GL, Zhu MB, Li Z, Yang JW, Wang ZQ. 2022. A prediction model of relativistic electrons at geostationary orbit using the EMD-LSTM network and geomagnetic indices. Space Weather. 20(10):e2022SW003126. doi: 10.1029/2022SW003126.

Zheng C, Kasprowicz CG, Saunders C. 2017. Customized routing optimization based on gradient boost regressor model. arXiv preprint arXiv:1710.11118.

Zou R, Yang Z, Zhang J, Lei R, Zhang W, Fnu F, Lei H, Heyne J, Zhang X, Ruan R. 2024. Machine learning application for predicting key properties of activated carbon produced from lignocellulosic biomass waste with chemical activation. Bioresource Technol. 399:130624. doi: 10.1016/j.biortech.2024.130624.