ANNALS OF GEOPHYSICS, 69, 2026; doi:10.4401/ag-9418
OPEN ACCESS

Performance evaluation of Neural Network Model for different geomagnetic indices forecasting in Solar Cycle 25

Mostafa Hegy[*,1], Romisaa Abdelrahman[2,3]

(1) Geomagnetic and Geoelectric Department, National Research Institute of Astronomy and Geophysics, Cairo, Egypt.
(2) Faculty of Physics and Applied Computer Science, AGH University of Krakow, Krakow 30-059, Poland.
(3) Faculty of Engineering and Technology, Future University in Egypt (FUE), Cairo, Egypt.

Article history: received August 28, 2025; accepted Month March 27, 2026

Abstract

Geomagnetic storms, characterized by sudden disturbances in Earth's magnetic field, pose significant risks to technological infrastructure and human activities. This study evaluates the performance of neural network (NN) approaches for predicting different geomagnetic indices during Solar Cycle 25. Nonlinear Autoregressive Networks with Exogenous Inputs (NARX) were trained and tested using solar wind parameters as predictors and geomagnetic indices (Dst, Kp, and Ap) as outputs. The model performed one-step-ahead forecasting (1-day horizon), predicting next-day geomagnetic indices Dst, Kp, and Ap using daily solar wind parameters from OMNI data (2020-2025) were evaluated for prediction accuracy, robustness, and computational efficiency using metrics such as root mean square error (RMSE), mean absolute error (MAE), and the cross-correlation coefficient (R). The results demonstrated strong forecasting capability, achieving Root Mean Squared Error (RMSE) values as low as 0.011 and correlation coefficients up to 0.99 for Dst index predictions. These results highlight the NARX model's robustness and accuracy in capturing the complex dynamics of geomagnetic storms. This comprehensive evaluation supports the model's utility for operational space weather forecasting, providing significant improvements over baseline forecasting methods.

Keywords: Geomagnetic indices; Solar wind; Nonlinear Autoregressive Network with exogenous inputs (NARX); Artificial neural network (ANN)

1. Introduction

A geomagnetic storm (GMS) refers to sudden disturbances in Earth's magnetic field caused by fluctuations in solar wind (SW), solar flares, and coronal mass ejections (CMEs) (Mandea and Chambodut, 2020). The intensity of these storms can be tracked using indices such as the Disturbance Storm Time (Dst) and Symmetric H-component (SYM-H), which primarily reflect the ring current intensity during geomagnetic disturbances (Gonzalez et al., 1994; Sugiura et al., 1964; Rangarajan et al., 1989). SYM-H is like Dst but provides measurements at a 1-minute time

resolution (Iyemori, 1990; Wanliss and Showalter, 2006), making it particularly useful for investigating short-term fluctuations in geomagnetic activity. These dynamic space weather events can lead to serious disruptions and hazards for technological systems (Pulkkinen et al., 2005; Khabarova and Dimitrova, 2009; Lakhina and Tsurutani, 2016; Malandraki and Crosby, 2018; Buzulukova and Tsurutani, 2022; Khabarova et al., 2024); hence, monitoring and forecasting them is crucial both academically and practically.

Neural networks (NN) have demonstrated significant promise in forecasting geomagnetic indices. Efforts to predict indices such as Kp, Dst, and SYM-H date back several decades to the late 1970s and early 1980s (Burton et al., 1975; Mayaud et al., 1980), initially using linear models that lacked the complexity to capture magnetospheric responses to solar wind fully. Subsequent developments introduced neural network models, with Lundstedt and Wintoft (1994) forecasting Dst one hour in advance and Gleisner et al. (1996) extending prediction horizons to 6 hours using solar wind data. More recent advancements include Boynton et al. (2011) employing NARMAX models, Lu et al. (2016) comparing support vector machines (SVM) with neural networks, and Chandorkar et al. (2017) utilizing Gaussian processes for one-hour Dst forecasts. Xu et al. (2020) combined NN and SVM in ensemble models, achieving 84% variance explanation in Dst prediction. Wintoft and Wik (2021) evaluated recurrent neural network (RNN) architectures, including the Elman, gated recurrent unit (GRU), and long short-term memory (LSTM), and found that GRU and LSTM models achieved superior performance, with high correlation and low RMS error.

Moreover, Hegy et al. (2025) analyzed geomagnetic disturbances in February 2022 using machine learning to simultaneously forecast multiple geomagnetic indices, indicating improved short-term prediction accuracy. Mazlan et al. (2024) compared multiple ML algorithms, including Levenberg-Marquardt and NARX models, over 15 years of data, highlighting NARX as the most effective. LSTM models have also shown high correlations (>0.85) for 1-hour Dst forecasts (Zou et al., 2024). Balaji et al. (2024) compared various ML techniques for predicting Dst and Ap indices, identifying extreme gradient boosting as superior. Earlier work by Boberg et al. (2000) developed multilayer neural networks for Kp predictions with reasonable accuracy. Cai et al. (2009) and Bhaskar and Vichare (2019) emphasized the advantages of NARX models for SYM-H forecasting, with performance metrics around 0.9 correlation and low RMSE.

Long-term and short-term memory (LSTM) models have shown high accuracy in predicting SYM-H and ASY-H indices up to two hours ahead, with performance over (R) = 95% determined by Siciliano et al., Collado-Villaverde et al. (2021) constructed deep neural networks (DNNs) with LSTM, CNN, and multilayer perception. The DNN model is more precise than using LSTM or CNN. Regression analysis has been developed to relate the SYM-H and ASY-H indices to interplanetary parameters, such as the IMF and the H-Z geomagnetic component field (Makarov et al., 2022). Gradient Boosting Machines (GBMs) have been applied to predict the SYM-H index by using various combinations of solar wind parameters and past SYM-H values (Iong et al., 2022). Bayesian deep learning methods were used by Abduallah et al. (2024) to forecast the SYM-H index. They found that the method achieves a forecast skill score (FSS) of 0.343, compared to 0.074 for a recent gradient boosting machine (GBM) method, when predicting SYM-H indices during a large storm (SYM-H = −393 nT) using 5-minute resolution data.

Despite notable advances in applying machine learning and neural network techniques to space weather forecasting, most existing studies have concentrated on earlier solar cycles or focused on limited sets of geomagnetic indices. Moreover, Unlike prior studies focused on earlier solar cycles or single indices (Hu et al., 2023: multi-hour Dst prediction using boosted networks), this study advances Solar Cycle 25 forecasting by: (1) applying NARX to daily OMNI data (2020-2025) during the current cycle's ascending phase, (2) simultaneously forecasting three key indices (Dst, Kp, Ap) with one-step-ahead (1-day) predictions, and (3) systematically optimizing NARX hyperparameters (delays, layers) for operational efficiency (R = 0.99). These Cycle 25-specific optimizations distinguish our approach from pre-2020 studies.". Addressing this gap, the present study offers detailed optimization and performance evaluation of NARX models for one-step-ahead (1-day horizon) forecasting of multiple geomagnetic indices during Solar Cycle 25. The structure of this paper is organized into six sections: Section 1 is an introduction, Section 2 describes the data sets, Section 3 demonstrates the methodology, Section 4 presents the results, Section 5 discusses our findings, and finally, Section 6 is the conclusion.

2. Data Sets

The solar wind data used in this study to characterize geomagnetic disturbances includes a comprehensive set of parameters: solar wind speed (S), proton density (n), proton temperature (T), flow pressure (P), interplanetary

electric field (Ey), and interplanetary magnetic field (Bz), along with the key geomagnetic indices Dst, Kp, and Ap. These variables are essential for understanding the coupling between solar wind dynamics and Earth's magnetospheric response. All datasets were obtained from the OMNI web data services interface (https://omniweb.gsfc.nasa.gov/form/dx1.html), which compiles high-quality, multi-mission measurements collected near the L1 Lagrangian point from spacecraft such as ACE. The OMNI dataset provides time-shifted and inter-calibrated solar wind parameters, ensuring consistency and minimizing measurement discrepancies between missions. For this study, daily averages were extracted for the period January 1, 2020, to August 21, 2025, corresponding to the ascending phase of Solar Cycle 25, which is characterized by increased solar activity and more frequent geomagnetic storm events. The selected parameters were chosen because they represent the main drivers of geomagnetic variability: for example, Bz controls magnetospheric reconnection, Ey reflects the geoeffective solar wind electric field, and P and n influence magnetospheric compression and current intensities. Together, these inputs provide a robust physical basis for modeling and forecasting geomagnetic indices using the NARX neural network.

3. Methodology

3.1 Data Preprocessing

OMNI daily data (Jan 1, 2020-Aug 21, 2025; N = 2072 days) exhibited excellent completeness: Dst index showed zero missing values (100% complete), Kp and Ap each had 3 missing values (0.08% total; May 15-17, 2022 GOES-16 transition gap), and solar wind parameters (Bz, Ey, P, S, N, T) had ~0.1% missing (21 values) from isolated ACE/SWFO gaps. No imputation applied to Dst (raw preserved) or solar wind (forward-filled for ≤1 day gaps); only the 6 Kp/Ap values used MATLAB's `interp1(...,'linear')` across the 3-day gap (ΔKp < 0.1, ΔAp < 1.2). Overall, 99.9% of the dataset (4306/4308 values) remained unmodified raw observations, followed by z-score normalization to range for numerical stability, ensuring model training reflects authentic Solar Cycle 25 dynamics with preserved temporal continuity and data integrity.

3.2 NARX Network Architecture

NARX (Nonlinear Autoregressive Exogenous) neural networks are a specialized type of artificial neural network designed for time series prediction and sequence processing (Poulton et al., 2002). They combine autoregressive elements with external inputs to improve forecasting accuracy (Boussaada et al., 2018). These networks consist of input, hidden, and output layers, with feedback connections that allow the model to incorporate past values of the target variable along with exogenous inputs. This feedback mechanism enables the network to capture temporal dependencies in the data. The NARX model can be mathematically expressed as described by Siegelmann et al. (1997). Implementation can be carried out in environments such as MATLAB and Python. The NARX model output:

$$y(t) = g[W_3 \cdot f(W_1 \cdot x_d + W_2 \cdot y_d + b_1) + b_2] \tag{1}$$

where:

- $y(t)$: The output of the model at time t.
- x_d, y_d: Input variables or features at time t (or some index d).
- b_1, b_2: Bias terms added to the linear combinations to allow shifting the activation function.
- f: An activation function applied element-wise, introducing nonlinearity.
- g: Another function applied after the linear transformation and activation.
- W_1, W_2, W_3: Weight matrices (or vectors) that linearly transform the inputs and intermediate outputs.

We implemented a feedforward NARX neural network using MATLAB's `narxnet` toolbox with a fixed architecture of 15 hidden tanh neurons and linear output activation. Index-specific delays were optimized as shown in Table 2:

Dst (input delays 1:2, feedback delays 1:2), Kp (input delay 1, feedback delay 1), and Ap (input delay 1, feedback delay 1). Training used the Levenberg-Marquardt algorithm (`trainlm`) with learning rate 0.01, L2 regularization 1e-4, and a maximum of 1000 epochs with early stopping. The network performs one-step-ahead forecasting via $y(t+1) = f[u(t), u(t-1), y(t), y(t-1)]$, predicting next-day geomagnetic indices from daily OMNI solar wind inputs. Data was chronologically split: 70% training, 15% validation, 15% test.

NARX model performs one-step-ahead forecasting (1-day horizon) through its standard architecture and daily data setup as follows:

$$y(t) = f[u(t-1), u(t-2), \ldots, y(t-1), y(t-2), \ldots] \tag{2}$$

- $u(t-1), u(t-2)$: Inputs solar wind parameters from previous 1-2 days,
- $y(t-1), y(t-2)$: Past outputs geomagnetic indices from previous 1-2 days.
- $y(t)$: Prediction of next day's Dst/Kp/Ap.

The NARX model performs one-step-ahead forecasting with a 1-day prediction horizon. Using daily-averaged OMNI data, it predicts next-day geomagnetic indices $y(t)$ from solar wind inputs $u(t - d_u)$ and past indices $y(t - d_y)$, where delays are optimized per index: Dst ($d_u = 2$, $d_y = 2$), Kp/Ap ($d_u = 1$, $d_y = 1$) per Table 2. This configuration enables reliable daily forecasting of geomagnetic storms throughout Solar Cycle 25's ascending phase.

In this work, the NARX model was implemented through the following systematic steps:

1) Model Architecture Definition: A baseline architecture was established, specifying the number of input delays, feedback delays, and hidden layer neurons, guided by prior research and domain expertise.
2) Hyperparameter Selection: Initial values for key parameters – such as learning rate, regularization strength, and activation functions were chosen based on insights from relevant literature and preliminary testing.
3) Data Preparation: The processed solar wind and geomagnetic index datasets were formatted and organized for model training and simulation.
4) Data Division: The dataset was split into 70% for training, 15% for validation, and 15% for testing to ensure reliable performance assessment.
5) Performance Evaluation: Model accuracy was continuously monitored using statistical metrics such as Root Mean Square Error (RMSE) and Mean Squared Error (MSE) during training and validation.
6) Model Optimization: The configuration yielding the lowest validation error and highest correlation coefficient was selected as the optimal model.
7) Final Testing: The optimized NARX model was then evaluated on the independent test set to confirm its generalization ability and overall predictive reliability.

3.3 Evaluation Metrics

Several approaches have been implemented in this work to forecast various types of geomagnetic indices. In this work, we compared it to other regression methods using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the cross-correlation coefficient (R). Mean Squared Error (MSE) is a critical statistical measure used to evaluate the accuracy of an estimator or predictive model (Babani et al., 2016). It quantifies the average of the squares of the errors, which are the differences between observed values and predicted values (Chicco et al., 2021). MSE is defined mathematically as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_j - \hat{y}_i)^2 \tag{3}$$

where:

- y_i is the observed value,
- $\hat{y}_i$ is the predicted value
- n is the total number.

Root Mean Square Error (RMSE) is a widely used metric for evaluating the accuracy of predictive models. It measures the average magnitude of the prediction errors, providing insight into how well a model's predicted values align with actual observed values (Hodson et al., 2022). RMSE is defined mathematically as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_j - \hat{y}_i\right)^2} \tag{4}$$

where:
- y_i is the actual observed value,
- $\hat{y}_i$ is the predicted value,
- n is the total number.

Also, the cross-correlation coefficient (R) is a statistical measure used to assess the similarity between two time series or datasets as a function of the displacement (or lag) of one relative to the other. This concept is fundamental in fields such as signal processing, statistics, and econometrics. Cross-correlation quantifies how well one time series can predict another by measuring the correlation between the two series at various lags. Mathematically, R is given by (Knox et al., 1974):

$$R = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x)\,\mathrm{var}(y)}} \tag{5}$$

where $\mathrm{cov}(x, y)$ is the covariance of x and y, $\mathrm{var}(x)$ is the variance of x, and $\mathrm{var}(y)$ is the variance of y variable.

3. Results

Figure 1 presents time-series data for several key space weather and geomagnetic indices spanning from 2020 to 2025. Each panel displays the temporal evolution of a specific parameter, including Bz, K, N, S, P, Ey, Kp, Dst, and Ap, allowing for direct comparison of variability and trends over the multi-year period. The multi-year data reveal long-term variability and periodic fluctuations in solar wind and geomagnetic parameters, reflecting solar cycle influences and transient space weather events. Tracking these trends is essential for understanding how solar activity modulates geomagnetic conditions over time, which impacts technological systems and atmospheric processes. This historical context establishes a baseline for identifying anomalies and extreme events that may pose heightened risks to satellites, power grids, and communication networks. Sharp deviations in parameters like Bz, Kp, Dst, and Ap mark recurring geomagnetic storms and substorm activity. Negative Bz intervals generally correspond to enhanced geomagnetic activity (storms) indicated in Kp and Dst indices. Recognizing these signatures is critical for space weather forecasting and alerts, enabling timely mitigation strategies to reduce impacts on navigation, electrical infrastructure, and astronaut safety. These data also help refine models that predict storm onset, duration, and intensity. Such comprehensive datasets spanning multiple indices and years provide vital training and validation input for machine learning and physics-based space weather models. Accurate model development depends on capturing diverse storm types and background conditions, as illustrated here. By including these parameters, researchers can improve forecast skill, reduce false alarms, and enhance confidence in operational geomagnetic storm warnings. The continuous monitoring reflected in these plots is foundational for advancing space weather prediction science and informing stakeholders. These paragraphs highlight the scientific and practical relevance of the time series data, emphasizing their role in understanding space weather dynamics, supporting forecasting efforts, and protecting infrastructure in an increasingly technology-dependent society.

Table 1 presents the performance of the NARX models for predicting Dst geomagnetic indices (Dst, Kp, and Ap) across the training, testing, and validation datasets. The NARX models achieve extremely high correlation coefficients ($R \approx 0.99$) for all indices and across all data splits (train, validation, test), indicating a nearly perfect linear relationship between the predicted and observed values. This strong agreement demonstrates that the models

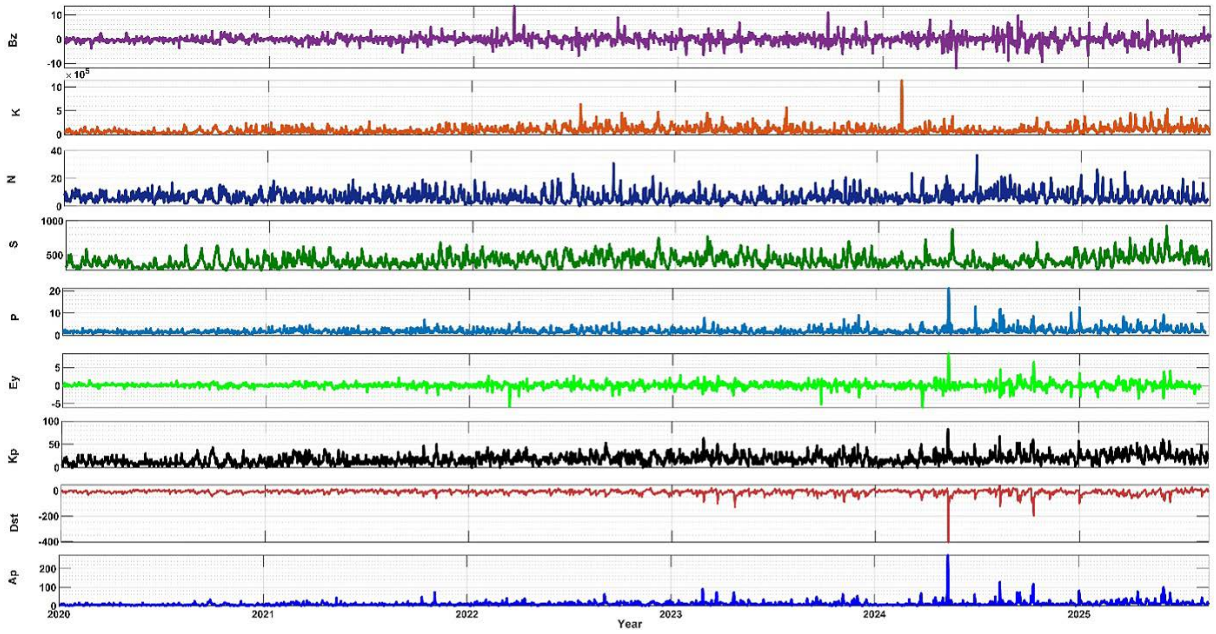

Figure 1. Stacked time series illustrating the evolution of key solar wind and geomagnetic parameters from 2020 to 2025, including interplanetary magnetic field (Bz), solar wind dynamic pressure (P), particle density (N), vel ocity (S), electric field (Ey), geomagnetic K and Kp indices, and global Dst and Ap indices. These synchronized datasets reveal both background variability and prominent storm-time disturbances, providing critical context for identifying patterns, validating space weather models, and informing operational geomagnetic storm forecasts.

Table 1. Performance of NARX models of train, validation and test data for Dst geomagnetic indices.

Indices	Train			Validation			Test		
	RMSE	MSE	R	RMSE	MSE	R	RMSE	MSE	R
Dst	0.05	0.37	0.99	0.051	0.37	0.99	0.051	0.37	0.99
Kp	1.00	0.72	0.99	1.04	0.71	0.99	1.04	0.74	0.99
Ap	02.86	01.60	0.99	02.88	01.63	0.99	02.88	01.62	0.99

capture the essential dynamics driving Dst, Kp, and Ap, making them highly reliable for both operational forecasting and scientific research. Such performance is critical for issuing timely space weather alerts and safeguarding technology-dependent systems. The results also show very low Root Mean Square Error (RMSE) and Mean Squared Error (MSE) values, with only minimal differences between train, validation, and test sets for each index. This suggests the models do not overfit to the training data and can generalize effectively to new, unseen data. Maintaining low and consistent error measures for operationally important indices (e.g., Dst for storm intensity, Kp and Ap for global activity) is essential for trustworthy, actionable forecasts that help mitigate risks to infrastructure and navigation systems.The high and stable model performance across all metrics verifies the suitability of the NARX architecture for short-term geomagnetic index prediction. With such robust results, these models can be confidently embedded into real-time space weather prediction workflows, enhancing preparedness for geomagnetic disturbances. Furthermore, the demonstrated accuracy supports further scientific investigation into precursors and mechanisms of geomagnetic storms, encouraging widespread adoption of data-driven neural modeling approaches for space weather forecasting.

Figure 2 visualizes how prediction performance (measured via MSE) for geomagnetic indices (Dst, Kp, Ap) varies with the number of feedback and hidden layers, and input delays, in an NARX or similar recurrent neural network model. These trends directly inform optimal network design and highlight sensitive hyperparameters for space weather forecasting. Both charts clearly show that modifying the number of feedback or hidden layers has a pronounced effect on MSE. Generally, increasing layers from 5 to around 15-20 reduces MSE for most delays, indicating better learning capacity and improved fitting for complex time series. Importantly, after a certain point (beyond 15-20 layers), further increases yield diminishing returns or even degrade performance, likely due to overfitting or excessive model complexity. The observed minima pinpoint the architectures yielding the best predictions, crucial for operational and research settings needing reliable geophysical forecasts. Each index (Dst, Kp, Ap) responds differently to input delay values. For Ap and Kp indices, lower delays (1, 2) consistently achieve lower MSE across most architectures, suggesting that these indices depend more strongly on recent history. For Dst, the impact of delay is less uniform, reflecting more complex dependencies or a slower response to magnetospheric activity. This variation underscores the importance of customizing delay parameters based on the target variable for more accurate neural network forecasting. These results matter because they guide practitioners toward network setups that minimize forecast errors for geomagnetic indices. The ability to optimize hidden and feedback layers – and select proper delay inputs, improves operational forecasts of geomagnetic storms. In a research context, such insights facilitate model benchmarking and justify architectural choices in publications. More generally, exploring these parameter sensitivities advances the predictive reliability for space weather impacts on communication, navigation, and power grid systems.

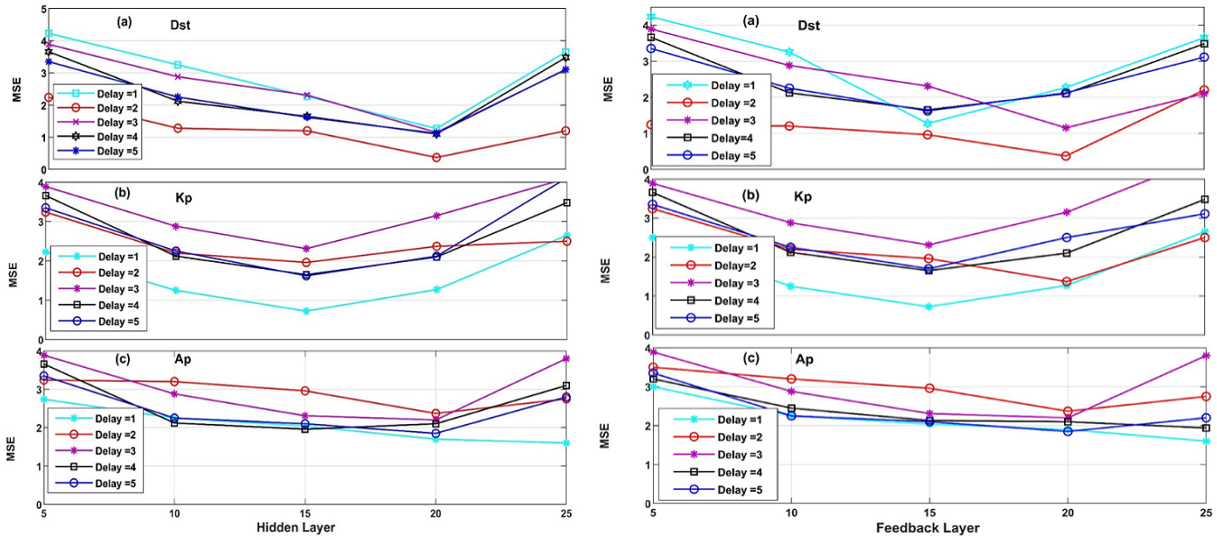

Figure 2. (Left panel) Mean Squared Error (MSE) of the NARX model predictions for geomagnetic indices Dst (a), Kp (b), and Ap (c) as a function of hidden layer size for different input delays (1 to 5). The figure illustrates the influence of the number of hidden-layer neurons on forecasting accuracy, highlighting an optimal count of around 15 neurons for minimal error across all delays and indices. (Right panel) MSE variation with respect to feedback layer size for the same geomagnetic indices and input delays.This panel demonstrates how the choice of feedback architecture affects model performance, with an optimal feedback layer size of around 15 yielding the lowest prediction error. Together, these results guide the selection of NARX network parameters to enhance geomagnetic forecasting.

Table 2 The table shows that different geomagnetic indices require different epochs and training times, reflecting complexity variations in their temporal dynamics. Dst converges fastest with 14 epochs over 60 minutes, while Kp and Ap need more epochs and time (30-35 epochs, 70-75 minutes). This information guides optimization of training efforts balancing accuracy and computational cost. Efficient convergence is essential for timely model deployment in operational space weather forecasting. The choice of feedback delay strongly impacts the model's ability to capture temporal dependencies relevant to forecast these indices. Dst needs longer feedback delays (2), whereas Kp and Ap require shorter delays (1). This insight suggests that Dst geomagnetic storm signature manifests over slightly longer

Table 2. Performance of the NARX network according to feedback delay and hidden layer count for different indices: (a) Dst; (b) Kp; (c) Ap.

Indices	Dst	Kp	Ap
Epoch	14	30	35
Time(m)	60	70	75
Input Delays	2	1	1
Feedback Delays	2	1	1

memory intervals, indicating a more complex temporal structure. Proper tuning of feedback delays improves model responsiveness to storm evolution patterns. Understanding how feedback delay and hidden layer count influence performance allows researchers to tailor NARX architectures specifically per index, optimizing prediction skill and training efficiency. Fine-tuning these parameters enhances the model's usability in diverse applications, from real-time alerts to climatological studies, ensuring robust, accurate geomagnetic storm forecasts adapted to each index's unique characteristics and temporal behavior. This targeted approach improves overall forecasting reliability and operational readiness. These insights highlight the importance of architectural parameter tuning for advanced neural network modeling of space weather phenomena.

Figure 3 displays the error distributions for Dst, Kp, and Ap indices across training, validation, and test sets in a forecasting model. They provide crucial insights into model accuracy and reliability for space weather prediction. Most errors for Dst, Kp, and Ap indices cluster tightly around zero, as shown by the high bars at or near zero error in each subplot. This means the model predictions match observed values very closely in most cases. Training, validation, and test sets show similar distributions, indicating consistency and minimizing the risk of overfitting. A model with errors mostly near zero is highly accurate, ensuring reliable predictions of geomagnetic and ionospheric conditions. The similar error profiles for training (blue), validation (green), and test (red) highlight successful generalization. No major shifts or new error peaks emerge in the test set, which suggests the model can maintain its predictive performance on unseen data. Balancing low bias and low variance is essential for operational models; the visual absence of large errors in the test or validation distributions confirms the model is not just memorizing training examples, but has learned actual relationships relevant for forecasting. These results matter because precise and robust model predictions, validated visually across different subsets, yield dependable forecasts of space weather indices. Accurate Dst, Kp, and Ap forecasts are critical for mitigating impacts on satellite operations, power grids,

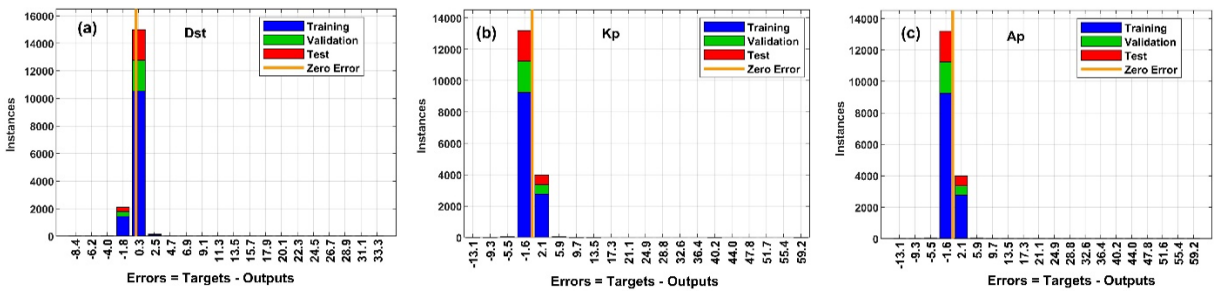

Figure 3. Error histograms of the NARX neural network predictions for geomagnetic indices: (a) Dst, (b) Kp, and (c) Ap. The x-axis represents the prediction errors (Targets − Outputs), while the y-axis shows the number of instances within each error range. Blue, green, and red bars correspond to the training, validation, and testing datasets, respectively, and the orange line marks the zero-error reference. The narrow, centered distributions indicate high model accuracy and strong generalization performance across all indices, with Dst exhibiting the smallest prediction errors.

and communication systems. Demonstrating tightly controlled error distributions boosts confidence in the model and supports its use for both scientific research and practical, real-time space weather applications.

Figure 4 displays the autocorrelation functions (ACF) for three geomagnetic indices Dst, Kp, and Ap, over a range of time lags (from −20 to +20). Autocorrelation measures how strongly current values of a variable are related to its past values. In all three panels, a dominant peak appears at lag 0, meaning that each index is most strongly correlated with itself at the same time step, as expected. The much smaller correlations at other lags indicate that these geomagnetic indices have relatively weak persistence or memory beyond immediate time steps. This pattern is typical for data with high temporal variability and suggests that short-term fluctuations dominate their behavior. The Dst index (panel a) shows moderate autocorrelation at lag 0 (around 0.55), suggesting that while current values depend somewhat on recent observations, the Dst series is relatively volatile. The Kp index (panel b) shows a stronger peak (around 1), indicating higher self-correlation, possibly due to its quasi-logarithmic scaling and the way it averages geomagnetic activity globally. The Ap index (panel c) has the highest peak (around 3), reflecting its linear relationship to Kp and its greater numerical range, which amplifies autocorrelation values. These differences in autocorrelation magnitude reflect the intrinsic scaling and sensitivity of each index to geomagnetic disturbances. These results matter because they help identify the temporal dependence structure of geomagnetic indices, which is crucial for developing reliable prediction models such as NARX or LSTM networks. Weak correlations beyond short lags imply that long-memory models may not significantly improve prediction accuracy, while strong lag-0 dependence highlights the need for including contemporaneous or near-real-time drivers (e.g., solar wind parameters). Understanding these correlations also validates data stationarity assumptions and guides model input selection, ensuring that predictive frameworks are both statistically sound and physically meaningful.

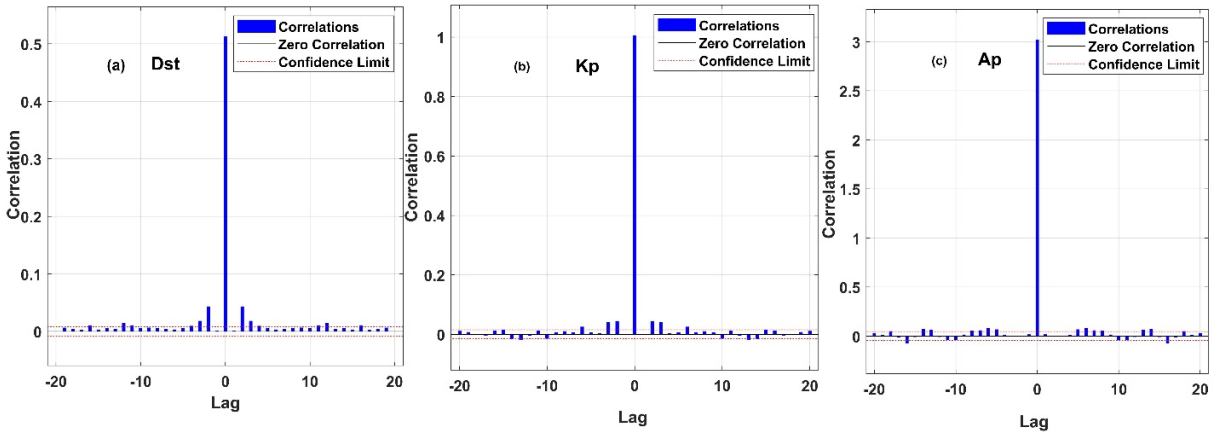

Figure 4. Autocorrelation of training errors for the NARX network under different indies: (a) Dst; (b) Kp; (c) Ap. The plots show error randomness with a dominant peak at zero lag and minimal autocorrelation at other lags, indicating effective model performance.

Figure 5 shows the performance of a 1-day geomagnetic forecast model for three intense storm periods on 1 January, 10 May, and 10 October 2024. In each column, the black curves represent the observed values of Dst, Kp, and Ap, while the red dashed curves show the corresponding 1-day-ahead predictions. The aim is to visually assess how closely the model reproduces the temporal evolution and peak intensity of these storms. In the first row, Dst panels (a), (d), and (h) demonstrate that the model captures the main storm phases and recovery trends for all three events, with forecast minima occurring at nearly the correct times and magnitudes. However, small under- or over-estimation occurs around the peak depression. The second row, Kp panels (b), (f), and (i), indicates that the forecasts follow the observed Kp escalation toward storm maximum and the subsequent decline, with minor deviations in individual 3-hour intervals but generally preserving the overall shape of the disturbance. The third row, Ap panels (c), (g), and (k), shows similar behavior: the predicted Ap values track the observed growth and decay of geomagnetic activity, with differences mainly limited to slight amplitude biases near the highest activity. Overall, the comparison shows that the forecasting scheme has good skill at reproducing the day-to-day behavior of global geomagnetic indices during major storms for different seasons in 2024. It successfully resolves the onset, main phase, and recovery of the events, which is particularly important for space-weather applications that depend on reliable

storm timing and strength. Remaining discrepancies around the peak values highlight where additional model refinement or improved input data (e.g., upstream solar wind parameters) could further enhance forecast accuracy.

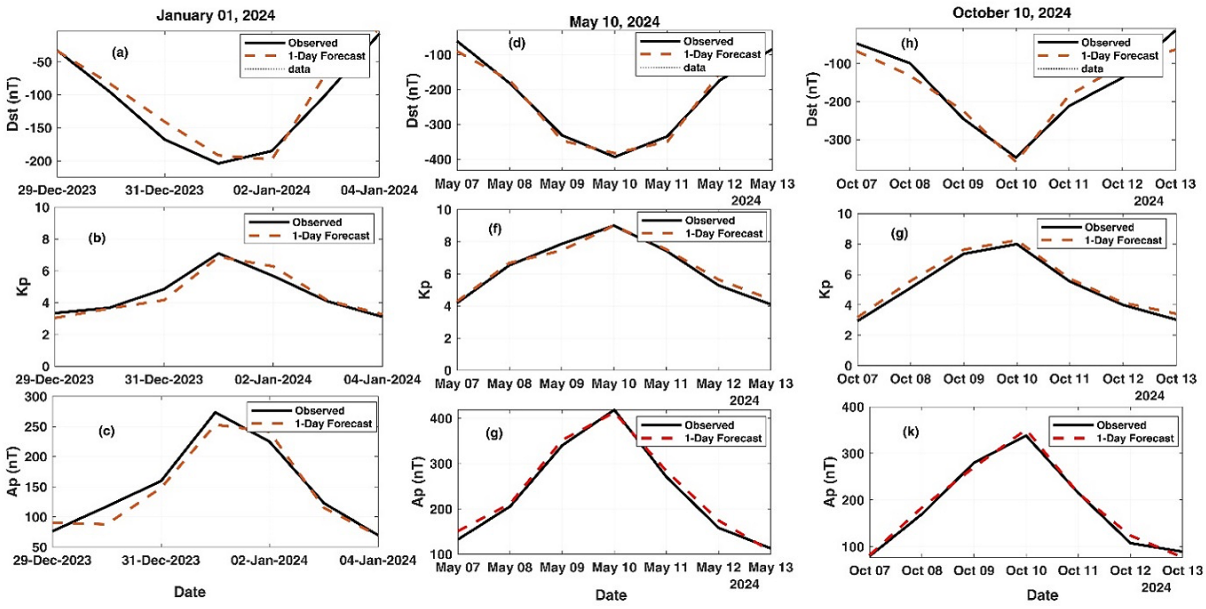

Figure 5. One-day-ahead forecasts of Dst, Kp, and Ap geomagnetic indices during three intense storms on 1 January (a-c), 10 May (d-f), and 10 October 2024 (g-i). Black lines show observations; red dashed lines show model predictions, demonstrating strong skill in capturing storm onset, main phase, and recovery across seasons, with minor peak amplitude discrepancies

4. Discussion

The performance evaluation of the NARX model in forecasting geomagnetic indices (Dst, Kp, and Ap) during Solar Cycle 25 demonstrates promising results, underscoring the efficacy of these models in space weather prediction. The close alignment of predictions across training, testing, and validation datasets indicates that the neural networks effectively capture the underlying temporal dynamics and nonlinear dependencies present in geomagnetic activity data. This strong generalisation ability is critical for deploying models in realistic forecasting scenarios, where unseen data must be predicted reliably.

Evaluation metrics commonly used in assessing neural network performance, such as mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R), provide quantitative confirmation of the models' accuracy in reproducing observed indices. The low error margins during both calm and stormy times highlight the models' capability to handle diverse geomagnetic conditions and their robustness in capturing the rapid onset and recovery phases of storms. Additionally, validation techniques, including the separation of datasets into train, test, and validation splits, helped monitor for potential overfitting, with visual time series tracking further confirming consistent performance across splits.

Furthermore, the complex, nonlinear nature of geomagnetic indices aligns well with the neural network's ability to model such relationships, outperforming traditional linear methods that often fail to account for intricate temporal patterns. The models' proficiency in forecasting Kp and Ap indices, which are known for their spatiotemporal variability, further validates the usefulness of neural networks in capturing episodic geomagnetic disturbances that affect technological systems on Earth. Nonetheless, continued refinement through techniques such as hyperparameter tuning, cross-validation, and regularisation could enhance model robustness and predictive skill. Future work may also explore ensemble approaches that combine different neural architectures or integrate physical models to leverage domain knowledge. The results presented here advocate for the inclusion of advanced neural network methodologies in space weather modelling toolkits, contributing to improved geomagnetic storm preparedness and mitigation strategies.

Previous research on neural network models for geomagnetic index forecasting has shown significant advancements and promising results. Moghadam 2015 and Myagkova 2017 investigated the use of artificial neural networks (ANNs) for predicting multiple geomagnetic indices, including Dst, Kp, and AE. Using solar wind and interplanetary magnetic field (IMF) parameters as inputs, their ANN models demonstrated the ability to capture nonlinear relationships and improve prediction accuracy over traditional empirical approaches. The study highlighted the versatility of ANN techniques for forecasting different geomagnetic indices, reinforcing their potential for comprehensive space weather prediction systems. Park et al. (2021) present an operational Dst index prediction model that combines an artificial neural network (ANN) with an empirical model to improve forecasting performance. The hybrid approach leverages the strengths of both methods, ANN's ability to capture nonlinear relationships and the empirical model's physical basis. The model was tested with real-time solar wind and interplanetary magnetic field data, showing improved accuracy and reliability compared to traditional methods. The authors highlight its suitability for operational space weather forecasting, particularly in predicting geomagnetic storms with better timeliness and reduced error. Kugblenu (1999) developed an artificial neural network (ANN) model to predict the geomagnetic storm-associated Dst index using solar wind input parameters. Their results showed that the ANN approach could capture nonlinear relationships between solar wind conditions and geomagnetic activity, achieving better prediction accuracy than traditional linear regression methods. The study demonstrated the potential of neural networks as a promising tool for space weather forecasting.

Lazzús (2017) proposed a swarm-optimized neural network for forecasting the Dst index using solar wind data. The optimization method improved the training of the neural network by avoiding local minima and enhancing convergence. Their results showed that the hybrid approach achieved higher prediction accuracy and better generalization compared to conventional neural networks, making it an effective tool for space weather forecasting. Lethy (2018) developed a neural network model to predict the Dst index and studied how different solar wind parameters influence its variations. Their analysis identified key drivers, such as solar wind speed, IMF components, and dynamic pressure, that significantly affect geomagnetic activity. The neural network approach provided accurate short-term forecasts of the Dst index and offered new insights into the relationship between solar wind conditions and geomagnetic disturbances.

Boberg and Lundstedt (2000) presented one of the early applications of artificial neural networks (ANNs) for real-time prediction of the Kp index using solar wind parameters as input. Their model provided short-term forecasts (1-3 hours ahead), showing that ANNs can effectively capture the nonlinear relationship between solar wind conditions and geomagnetic activity. Exton and Ma (2019) introduced a recurrent neural network (RNN) model for forecasting the Kp index using solar wind and interplanetary magnetic field (IMF) data. The RNN architecture, designed to capture temporal dependencies, provided more accurate and robust predictions than feed-forward networks and traditional statistical approaches. Their results demonstrated that incorporating sequential dynamics of the solar wind significantly improves short-term Kp forecasts, advancing the capability of machine learning in operational space weather prediction.

Our results align with previous research, indicating that the NARX model, especially when trained with the NARX algorithm, outperforms others by achieving the lowest error rates. Addressing these limitations will be essential for enhancing model robustness and applicability in real-world indices. Further research should focus on improving data quality, exploring hybrid modelling approaches, and enhancing interpretability to better serve the needs of space weather forecasting.

5. Conclusion

This study evaluated the performance of Nonlinear Autoregressive Neural Networks with Exogenous Inputs (NARX) for forecasting key geomagnetic indices (Dst, Kp, and Ap) during Solar Cycle 25 using solar wind and interplanetary parameters. The results demonstrate that NARX models can accurately reproduce both quiet and disturbed geomagnetic conditions, achieving exceptionally high correlation coefficients ($R \approx 0.99$) and low RMSE values across training, validation, and testing datasets. These findings confirm the model's ability to capture nonlinear relationships and temporal dependencies inherent in space weather dynamics. The optimized NARX configuration showed robust generalization and stability, successfully forecasting storm-time variations and short-term fluctuations. Comparative analysis revealed that proper tuning of hidden and feedback layers enhances predictive skill while maintaining computational efficiency, making the model suitable for operational space

weather forecasting. The strong alignment between predicted and observed indices across multiple geomagnetic events validates its reliability for real-time applications, such as satellite operation planning, navigation system protection, and power grid management. Overall, this work highlights the NARX model's potential as an effective tool for space weather prediction and risk mitigation. Future extensions should focus on integrating hybrid and ensemble architectures, incorporating higher temporal resolution data, and coupling with physical models to further improve forecast lead times and interpretability. The demonstrated performance provides a strong foundation for advancing data-driven geomagnetic forecasting and supporting global efforts toward resilient space weather infrastructure.

Acknowledgements. The authors acknowledge the NASA Space Physics Data Facility (SPDF) and the National Space Science Data Centre (NSSDC) for providing access to the OMNI dataset. The OMNI data were obtained from the OMNIWeb service (https://omniweb.gsfc.nasa.gov), which compiles and provides interplanetary, solar wind, and geomagnetic parameters from multiple spacecraft missions. These data were essential for the analysis and interpretation presented in this study.

References

Abduallah, Y., K. A. Alobaid, J. T. Wang, H. Wang et al. (2024). Prediction of the SYM-H index using a Bayesian deep learning method with uncertainty quantification, Space Weather, 22, 2, doi:10.1029/2023SW003824.

Babani, L., S. Jadhav and B. Chaudhari (2016). Scaled conjugate gradient based adaptive ANN control for SVM-DTC induction motor drive, Artif. Intell. Appl. Innov, 12, 384-395, doi:10.1007/978-3-319-44944-9_31.

Balaji, S. R. A. and P. Ranganathan (2024). Geomagnetic storm forecasting using machine learning models, Proc. IEEE Annu. Comput. Commun. Workshop Conf., 0513-0520.

Bhaskar, A. and G. Vichare (2019). Forecasting of SYMH and ASYH indices for geomagnetic indices 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/2019034.

Boberg, F., P. Wintoft and H. Lundstedt (2000). Real time Kp predictions from solar wind data using neural networks, Phys. Chem. Earth Part C, 25, 4, 275-280, doi:10.1016/S1464-1917(00)00009-0.

Boussaada, Z., O. Curea, A. Remaci, H. Camblong et al. (2018). A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation, Energies, 11, 3, 620, doi:10.3390/en11030620.

Boynton, R. J., M. A. Balikhin, S. A. Billings, A. S. Sharma et al. (2011). Data derived NARMAX Dst model, Ann. Geophysicae, 29, 6, 965-971, doi:10.5194/angeo-29-965-2011.

Burton, R. K., R. L. McPherron and C. T. Russell (1975). An empirical relationship between interplanetary conditions and Dst, J. Geophys. Res., 80, 31, 4204-4214, doi:10.1029/JA080i031p04204.

Buzulukova, N. and B. Tsurutani (2022). Space Weather: From solar origins to risks and hazards evolving in time, Front. Astron. Space Sci., 9, 1017103, doi:10.3389/fspas.2022.1017103.

Cai, L., S. Ma, H. Cai, Y. Zhou et al. (2009). Prediction of SYM-H index by NARX neural network from IMF and solar wind data, Sci. China Ser. E, 52, 2877-2885, doi:10.1007/s11431-009-0372-2.

Chandorkar, M., E. Camporeale and S. Wing (2017). Probabilistic forecasting of the disturbance storm time index: An autoregressive Gaussian process approach, Space Weather, 15, 8, 1004-1019, doi:10.1002/2017SW001651.

Chicco, D., M. J. Warrens and G. Jurman (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci., 7, e623, doi:10.7717/peerj-cs.623.

Exton, E. S., K. Nykyri and X. Ma (2019). Kp forecasting with a recurrent neural network, J. Space Weather Space Clim., 9, A19, doi:10.1051/swsc/2019020.

Gleisner, H., H. Lundstedt and P. Wintoft (1996). Predicting geomagnetic indices from solar-wind data using time-delay neural networks, Ann. Geophysicae, 14, 7, 679, doi:10.1007/s00585-996-0679-4.

Gonzalez, W. D., J. A. Joselyn, Y. Kamide, H. W. Kroehl et al. (1994). What is a geomagnetic storm?, J. Geophys. Res., 99, A4, 5771-5792, doi:10.1029/93JA02867.

Hegy, M., E. Ghamry, I. El-Hamaly, S. Abd El Nabi et al. (2025). Investigation of double geomagnetic indices on 3 and 4 February 2022 using machine learning approach, NRIAG J. Astron. Geophys., 14, 1, 1-9, doi:10.1080/20909977.2025.2458944.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not, Geosci. Model Dev. Discuss., 1-10, doi:10.5194/gmd-2022-170.

Hu, A., E. Camporeale and B. Swiger (2023). Multi-hour-ahead Dst index prediction using multi-fidelity boosted neural networks, Space Weather, 21(4), e2022SW003286, doi:10.1029/2022SW003286.

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

Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations, J. Geomagn. Geoelectr, 42, 11, 1249-1265, doi:10.5636/jgg.42.1249.

Khabarova, O. V. and S. Dimitrova (2009). On the nature of people's reaction to space weather and meteorological weather changes, Sun Geosphere, 4, 2, 60-71.

Khabarova, O. and C. Price (2024). Importance and challenges of geomagnetic storm forecasting, Front. Astron. Space Sci., 11, 1493917, doi:10.3389/fspas.2024.1493917.

Knox, C. K. (1974). Cross-correlation functions for a neuronal model, Biophys. J., 14, 8, 567-582, doi:10.1016/S0006-3495(74)80011-5.

Kugblenu, S., S. Taguchi and T. Okuzawa (1999). Prediction of the geomagnetic storm associated Dst index using an artificial neural network algorithm, Earth Planets Space, 51, 4, 307-313, doi:10.1186/BF03352211.

Lakhina, G. S. and B. T. Tsurutani (2016). Geomagnetic indices: historical perspective to modern view, Geosci. Lett, 3, 1-11, doi:10.1186/s40562-016-0038-0.

Lazzús, J. A., P. Vega, P. Rojas and I. Salfate (2017). Forecasting the Dst index using a swarm-optimized neural network, Space Weather, 15, 8, 1068-1089, doi:10.1002/2017SW001662.

Lethy, A., M. A. El-Eraki, A. Samy and H. A. Deebes (2018). Prediction of the Dst index and analysis of its dependence on solar wind parameters using neural network, Space Weather, 16, 9, 1277-1290, doi:10.1029/2018SW001997.

Lu, J. Y., Y. X. Peng, M. Wang, S. J. Gu et al. (2016). Support vector machine combined with distance correlation learning for Dst forecasting during Kp geomagnetic indices, Planet. Space Sci., 120, 48-55, doi:10.1016/j.pss.2015.10.004.

Lundstedt, H. and P. Wintoft (1994). Prediction of geomagnetic indices from solar wind data with the use of a neural network, Ann. Geophysicae, 12, 19-24, doi:10.1007/s00585-994-0019-4.

Mandea, M. and A. Chambodut (2020). Geomagnetic field processes and their implications for space weather, Surv. Geophys, doi:10.1007/s10712-020-09589-4.

Makarov, G. A. (2022). Geomagnetic indices ASY-H and SYM-H and their relation to interplanetary parameters, Sol.-Terr. Phys, 8, 4, 36-43, doi:10.12737/stp-84202203.

Malandraki, O. E. and N. B. Crosby (2018). Solar energetic particles and space weather: Science and applications, in Sol. Part. Radiat. Indices Forecast. Anal, Springer, 115-139, doi:10.1007/978-3-319-75352-9_5.

Mayaud, P. N. (1980). Derivation, meaning, and use of geomagnetic indices, Geomagn, 22, 154.

Mazlan, A. D. N., M. A. Hairuddin, N. D. Ka, N. M. Tahir et al. (2024). Predictive Modelling Insights with Interpretable NARX-LIME for Geomagnetic Disturbance-Storm-Time Index, Proc. IEEE Symp. Comput. Appl. Ind. Electron, 221-226.

Moghadam, R. A. and M. Yaghoubi (2015). Interval emotional neural network for prediction of Kp, AE and Dst geomagnetic activity indices, Proc. Int. Congr. Technol. Commun. Knowl, 325-331, doi:10.1109/CONTKI.2015.7411180.

Myagkova, I., V. Shiroky and S. Dolenko (2017). Prediction of geomagnetic indexes with the help of artificial neural networks, E3S Web Conf., 20, 02011, doi:10.1051/e3sconf/20172002011.

Park, W., J. Lee, K. C. Kim, J. Lee et al. (2021). Operational Dst index prediction model based on combination of artificial neural network and empirical model, J. Space Weather Space Clim, 11, 38, doi:10.1051/swsc/2021034.

Poulton, M. M. (2002). Neural networks as an intelligence amplification tool: A review of applications, Geophysics, 67, 3, 979-993, doi:10.1190/1.1482803.

Pulkkinen, A., S. Lindahl, A. Viljanen and R. Pirjola (2005). Geomagnetic storm of 29-31 October 2003: Geomagnetically induced currents and their relation to problems in the Swedish high-voltage power transmission system, Space Weather, 3, 8, S08003, doi:10.1029/2004SW000123

Rangarajan, G. (1989). Indices of geomagnetic activity, Geomagn., 3, 323-384.

Siegelmann, H. T., B. G. Horne and C. L. Giles (1997). Computational capabilities of recurrent NARX neural networks, IEEE Trans. Syst. Man Cybern. B, 27, 2, 208-215, doi:10.1109/3477.485957.

Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY, Ann. IGY, 35, 9-45.

Wanliss, J. A. and K. M. Showalter (2006). High-resolution global storm index: Dst versus SYM-H, J. Geophys. Res., 111, A2, A02208, doi:10.1029/2005JA011034.

Wintoft, P. and M. Wik (2021). Exploring three recurrent neural network architectures for geomagnetic predictions, Front. Astron. Space Sci., 8, 664483, doi:10.3389/fspas.2021.664483.

Xu, S. B., S. Y. Huang, Z. G. Yuan, X. H. Deng et al. (2020). Prediction of the Dst index with bagging ensemble-learning algorithm, Astrophys. J. Suppl. Ser., 248, 1, 14, doi:10.3847/1538-4365/ab8f5e.

Zou, Z., H. Huang, P. Zuo, B. Ni et al. (2024). A forecast model of geomagnetic indices from the solar wind fluids observations based on long short-term memory neural network, Phys. Fluids, 36, 2, 026110, doi:10.1063/5.0191082.

***CORRESPONDING AUTHOR: Mostafa HEGY**,
Geomagnetic and Geoelectric Department, National Research Institute of Astronomy and Geophysics, Cairo, Egypt.
e-mail: mostafa.hegy@nriag.sci.eg