ICIC Express Letters
Volume **19**, Number **5**, May **2025**

ICIC International ©2025 ISSN 1881-803X
pp. **485–496**

INTERPRETABLE N-BEATS DEEP NETWORKS OF MULTISTEP FORECASTING FOR THE GROUND-BASED GEOMAGNETIC DST INDEX

Nur Dalila Khirul Ashar[1,2], Syamsiah Mashohor[1,*], Aduwati Sali[3,4]
Mohamad Huzaimy Jusoh[5], Akimasa Yoshikawa[6]
Zatul Iffah Abdul Latiff[2] and Muhammad Asraf Hairuddin[2,7]

[1]Computer and Communication Systems
Faculty of Engineering
[3]WiPNET Department of Computer and Communication Systems
[4]Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia
Serdang 43400, Selangor, Malaysia
aduwati@upm.edu.my
*Corresponding author: syamsiah@upm.edu.my

[2]College of Engineering
Universiti Teknologi MARA Johor Branch
Jalan Purnama, Bandar Seri Alam 81750, Masai Johor, Malaysia
{ nurdalila306; zatul0130 }@uitm.edu.my

[5]School of Electrical Engineering
College of Engineering
[7]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI)
Universiti Teknologi MARA
Shah Alam 40450, Selangor, Malaysia
{ huzaimy; masraf }@uitm.edu.my

[6]International Research Center for Space and Planetary Environmental Science (i-SPES)
Kyushu University
744, Motooka, Nishi-ku, Fukuoka 819-0395, Japan
yoshikawa.akimasa.254@m.kyushu-u.ac.jp

Received June 2024; accepted September 2024

Abstract. *The importance of geomagnetic disturbances represented by ground earth activities based on the disturbance storm time (Dst index) entails an early forecast of geomagnetic storm occurrence, which could potentially disrupt the system operations. Often, the forecast outcome serves as an essential indicator for operational users who not only require early forecasting prior to incoming geomagnetic storms but also intend to obtain explainable insight and understanding of the generated forecast results. Therefore, a new model architecture, namely neural-basis expansion analysis for interpretable time series (N-BEATS), is proposed that incorporates a more transparent architecture of the deep learning model into producing the multiple steps ahead forecasting of the Dst index. Extensive comparisons among several deep learning models, namely long short-term memory (LSTM), gated recurrent units (GRU), and bidirectional GRU (Bi-GRU) network architectures, will be assessed, considering the model performances, and the impact of forecast variability will be discussed. The superiority of N-BEATS overcomes the state-of-the-art LSTM forecast model in terms of computational resources, and the effectiveness of learning the data of the Dst index pattern could be observed.*

Keywords: Neural beats model, Interpretable architecture, Disturbance storm time forecasting, Deep learning, Space weather

DOI: 10.24507/icicel.19.05.485

1. **Introduction.** Geomagnetic disturbance storm time (Dst index) forecasting poses significant hurdles to developing an accurate model that alleviates uncertainties in the developed model. This is due to the fact that complex nonlinear dynamics of interactions between geomagnetic storms and ground earth are challenging to capture [1,2]. Existing works, including empirical, statistical, and machine learning, have demonstrated the capabilities of various models, were examined through various means of prediction modelling development, and further expanded into forecast applications. The evolution of these methods somehow demonstrated positive improvements. In addition to huge data availability, the continuing progress has been remarkable for the Dst index forecasting.

Prior to the breakthrough in deep learning, neural networks were the most preferential use, which are composed of neurons and weights that learn and reduces the error from the input data to be mapped into the designated output [3-5]. In spite of the capabilities of neural networks demonstrated in capturing the nonlinearity complexity between data, network tuning to search for the best hyperparameters must be configured for the most optimal performance. Most recent approaches have prioritized the implementation of deep network architecture to improve geomagnetic Dst index forecasting. The deep learning algorithms possess advantages in capturing the complex dynamics in the dataset, recent emerging implementations have progressively been under development. Notably, the success of deep learning, which has been proven in several studies in the form of architectural advances and new frameworks, has been astonishing. Furthermore, it opens new avenues for efficient time-series forecasting to advance and outperform across a wide range of domains. Consequently, pioneering deep learning implementation into Dst index forecasting works rose, subsequently leading to the success of machine learning studies.

The present study's contribution focuses on the comparative analysis of different architectures of deep network models. The effectiveness of the forecasting model focusing on the geomagnetic disturbance index has been evaluated. However, no comparative evaluation exists for depth evaluation between various deep network architectures. A state-of-the-art model, namely N-BEATS, incorporates interpretability aspects in the architecture, improving multistep forecasting to attain the best performances for a longer forecast horizon for the Dst index. Inclusive interpretability assessments are essential for exploring new architectures by incorporating trends and seasonality into the model. Accordingly, this geomagnetic disturbance index model could be utilized as an operational forecast complement with fast computational results by considering long-term and complex data patterns into consideration. This is particularly useful for geophysics and space weather studies to allow proper mitigation actions involving earth technology systems. Therefore, it is significant in geomagnetic activity monitoring that early anticipates the future value for geomagnetic storm conditions.

The rest of the article's content is organized as follows. Section 2 provides the literature review related to the progress of geomagnetic index forecasting and prediction. Section 3 presents a detailed description of the workflow proposed architecture with various methods of deep networks-based forecasting techniques using LSTM, GRU, Bi-GRU, and N-BEATS. Section 4 discusses the simulation results and forecasting performance with point-to-point error metrics evaluation. In this section, the best-selected model will be proposed as a benchmark model for the geomagnetic Dst index. Section 5 provides the conclusions and future research directions.

2. **Related Research.** Several studies have been conducted to continuously improve the forecast model for use in geomagnetic activity monitoring, specifically the Dst index. Gruet et al. [6] initiated an LSTM model that could produce accurate predictions for up to three hours. The authors introduced the LSTM model with dense layers to form a Gaussian-type probabilistic forecast that successfully improved the forecast performance to a six-hour horizon at a correlation and RMSE of 87.3% and 9.86, respectively.

Additionally, several studies have examined the capabilities of Dst index predictions with a single architecture-based LSTM [7,8], modified LSTM networks [9-11], and a combination of LSTM architecture configurations with other networks [12]. For instance, Efitorov et al. [13] proposed a recurrent LSTM with 96 and 57 consecutive blocks with an Adam training gradient, 0.45 of dropout, and regularization settings. The forecast of twelve hours ahead demonstrated that the demerited performance of the multiple determination coefficient was 0.57, and the correlation coefficient was 0.76. Tasistro-Hart et al. [14] proposed a similar deep learning structure that measures the uncertainty. However, the authors highlighted that including regularizing terms is far more important to improve model reliability further. The farthest correlation coefficient reached up to six hours at 88.3%, which was reduced by 10.45% from the one-hour forecast by Wintoft and Wik [7] further explored the Dst forecast using three distinct recurrent neural network (RNN), namely GRU, Elman and LSTM. Comparable performances can be seen between GRU and LSTM, which outperform the Elman-based RNN model. The difference between models can be seen from the different architectures used in which the GRU and LSTM are composed of gating units that control the flow of information. Nevertheless, the GRU network has a less complex architecture, thus producing a shorter training time. The best model produced linear correlation and root mean square, respectively, at 0.90 and 8.8nT.

Other works have been proposed by improvising the network model. Abduallah et al. [15] attempted to forecast the Dst index using the Dst Transformer model, which combines a multi-head attention layer with Bayesian inference. The longest-horizon forecast results in R-squared and root mean square errors of approximately 60% and 5.0nT. Xu et al. [16] developed improved modelling using Bagging ensemble learning, which incorporates the combination of an LSTM network, artificial neural network, and support vector regression able to produce a six-hour forecast that demonstrates the root mean square error and correlation coefficient of 8.09nT and 85.72%. Cristoforetti et al. [17] introduced the most recent structure of using the LSTM with fully connected neural network (FC-NN). The model is composed of a modified LSTM network that concatenates the hidden states, which are further fed as input to the fully connected layer. The attempt to forecast the Dst index for up to twelve hours has demonstrated inconsistency in performance, which varies according to the storm phases and feed training data, securing at least 77% accuracy and 16% classification scores. Li et al. [18] implemented an improved LSTM model to include layer-wise relevance propagation (LRP) as a feature selector by decomposing the model prediction in the form of a relevance score and then further analyzing which features provided significant contributions.

The abovementioned works demonstrate impressive progress in forecasting the geomagnetic Dst index with various attempts on deep network models. While exploring deep networks, LSTM serves as the benchmark model, which contains a gating layer to retain past data. The model works well, especially for long data sets. However, it requires high computational processing power. The effectiveness of the forecasting model focusing on the geomagnetic disturbance index derived from geomagnetic observation data was evaluated. Nevertheless, the possibility of more interpretable techniques to compare various deep network architecture multi-step-ahead tasks has not been extensively explored to examine the deterioration in the reduced model performance. In fact, generalization variabilities exist at various levels of geomagnetic storm categories that could influence the performances of the deep learning model even though it can explore the hidden pattern in the dataset. The capability of deep learning by some means has improved over conventional Dst index prediction and forecasting. Hence, achieving a feasible approach to enhancing the Dst index into a long-term forecast horizon with high accuracy and interpretability of trustworthiness has been challenging.

3. Methodology.

3.1. Deep learning forecasting strategies for the Dst index. The main aim is to predict the future Dst index in the hourly resolution of data by considering the points forecasting approach from the past Dst index input. This paper proposes to compare the model performances among the different deep network architecture configurations as the benchmark model to forecast the univariate data of the Dst index. This includes the neural-basis expansion analysis for interpretable time series (N-BEATS), long short-term memory (LSTM), gated recurrent units (GRU), and bidirectional GRU (Bi-GRU). In this study, the Lazzús et al.'s model [19] functions as a base model reference with an improvement and extension in model architecture, and the dataset used was incorporated into the model. Model structure based on deep networks involves normalization and does not require an extensive pre-processing, trend, and seasonality analysis as per conventional time series forecasting analysis. Data normalization rescales input data to a distribution value between -1 and 1 with the aim that they can fit within the interval to result in faster convergence of gradient descent in minimizing weight and bias. Meanwhile, model performance compares how well the forecast and actual Dst index are performing. It is evaluated with several indicators' metrics relevant to error and correlation analysis to observe model variability. This comprises mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), mean absolute scaled error (MASE), training time, testing time, correlation coefficient (CC) and multiple coefficients of determination (R^2). Furthermore, elaborations on principle, network architecture and corresponding layers involved in the modelling are described in the next section.

3.2. N-BEATS. Newly designed architecture based on deep residual networks proposed by Oreshkin et al. [20] comprises several blocks, as shown in Figure 1. This consists of stacks of input (x_l) formed of a backcast link (θ_l^b), forecast link (θ_l^f), and four fully connected (FC) layers ($h_{l,1}, h_{l,2}, h_{l,3}, h_{l,4}$) with rectified linear unit (ReLU) activation function. The process starts at block input, which considers the estimated data after considering the lookback and future values. Then, the block takes these coefficients of θ_l^b and θ_l^f to produce the backcast ($\hat{x}_l$) and forecast ($\hat{y}_l$) signals. In the stack, these signals are

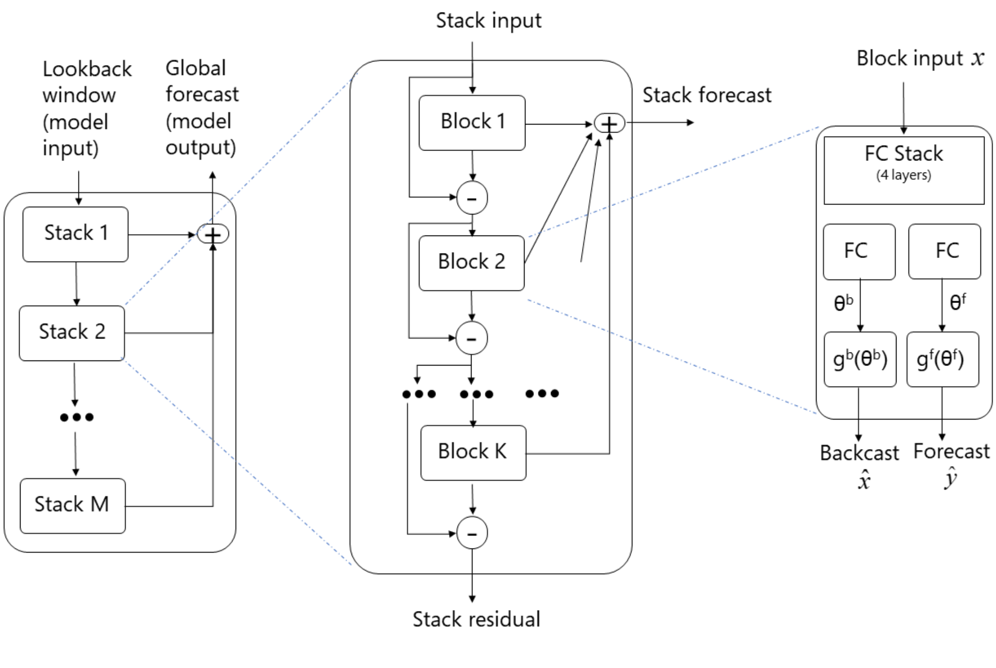


FIGURE 1. N-BEATS architecture

split into two separate branches in which each block subtracts the backcast signal from the input. Residuals are generated by subtracting the current value and looking back at data to feed further into the stack. Residual errors are then modelled from the previous block to improve forecast value and continuously repeat the process. This produces the final forecast $\hat{y}$, which accumulates all the forecast signals. Furthermore, the blocks are stacked in multiple forms to exploit the functionality similar to ensemble learning and avoid gradient vanishing issues.

This novel model, also known as pure deep learning, demonstrates high confidence in model performance in performing univariate, point-to-point, and multistep forecasts without specific expert knowledge. Notably, the proposed model has several known advantages, such as being fast to train and flexible to implement without modification.

$$h_{l,1} = FC_{l,1}(x_l), \quad h_{l,2} = FC_{l,2}(h_{l,1}), \quad h_{l,3} = FC_{l,3}(h_{l,2}), \quad h_{l,4} = FC_{l,4}(h_{l,3}) \tag{1}$$

$$\theta_l^f = LINEAR_l^b(h_{l,4}), \quad \theta_l^b = LINEAR_l^f(h_{l,4}) \tag{2}$$

$$\hat{y}_l = g_l^f\left(\theta_l^f\right), \quad \hat{x} = g_l^b\left(\theta_l^b\right) \tag{3}$$

$$x_l = x_{l-1} - \hat{x}_{l-1}, \quad \hat{y} = \sum_l \hat{y}_l. \tag{4}$$

3.3. **LSTM.** A network comprises memory cells and gating mechanisms that control the flow of information and learning long-term dependencies in the form of a data sequence. Specifically, the LSTM model is composed of the cell state and three gates, namely the input (i_t), output (c_t), and forget (f_t) gates, in which the cell transports the information while this information is either added or removed via these nonlinear gates (see Figure 2). These gates are essential for learning the information fed through the cell state and further keeping or forgetting the information during training. Notations for all parameters include x_t, o_t, h_t, t, i_t, $\hat{c}_t$, c_t, $w_f = w_c = w_o$, $b_f = b_c = b_o$, which present the input value, output value, hidden layer state value, moment, information that needs to be updated, cell state, candidate memory, weights, and biases.

$$\begin{aligned} c(t) &= f_t \odot c_{t-1} + i_t \odot \hat{c}_t \\ &= \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \odot c_{t-1} + \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \\ &\quad \odot \tanh\left(w_c \cdot [h_{t-1}, x_t] + b_c\right) \end{aligned} \tag{5}$$

$$o_t = \sigma\left(w_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{6}$$

$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

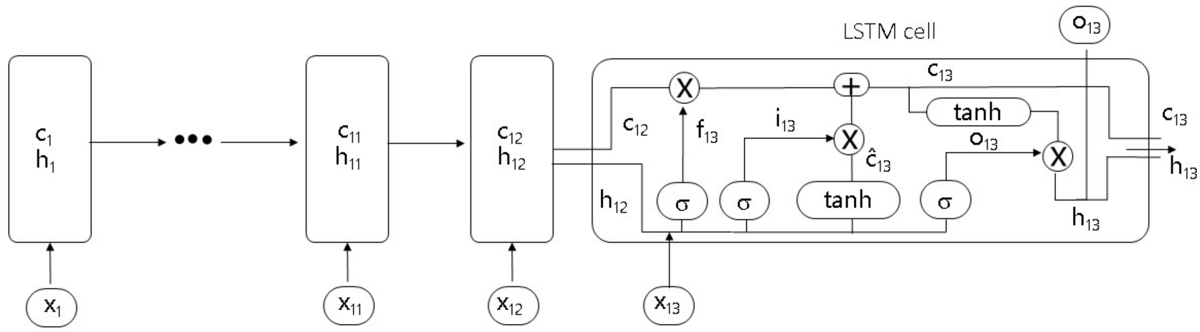


FIGURE 2. LSTM architecture

3.4. **GRU.** It is an improved version of the LSTM network, which utilized only two gate mechanisms, namely, reset gate (r_t) and update gate (z_t) (see Figure 3). In the update gate, the data are determined whether to be thrown or kept, while the reset gate decides the past information to forget. These gates perform operations without separate memory

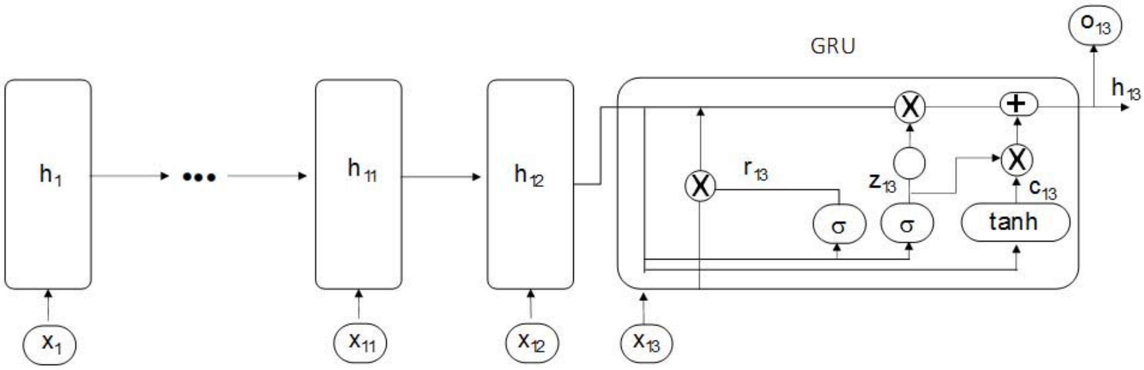


FIGURE 3. GRU architecture

to handle information flow. This memory works to retain, filter, and combine information from the previous state (h_{t-1}), to produce the current state (h_t), thereby reducing the architecture complexity of the networks, less tensor operations, and faster to train.

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t] + b_z) \tag{8}$$

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t] + b_r) \tag{9}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tanh(w_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \tag{10}$$

3.5. **Bidirectional GRU.** Further modification to improve the GRU working principle in unidirectional mode is by introducing the Bi-GRU to capture the information in dual directions, i.e., left-to-right (forward) and right-to-left (backward) directions (see Figure 4). Here, x_t, o_t, h_t, h_k respectively present input value, output value, hidden state of forward layer and hidden state of backward layer.

$$h_t = f(w_1 x_t + w_2 h_{t-1} + b_t) \tag{11}$$

$$h_k = f(w_3 x_t + w_4 h_{k-1} + b_k) \tag{12}$$

$$o_t = (h_t, h_k) \tag{13}$$

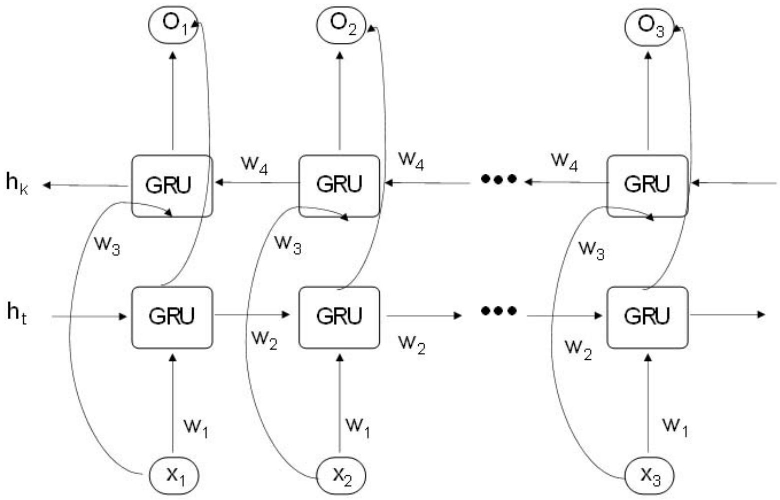


FIGURE 4. Bidirectional GRU architecture

This model consists of bidirectional connections that improvise the GRU architecture, allowing the model to simultaneously learn from both directions in a sequence of past and future data. With this extension form of GRU networks, the change in the Dst index in the past and future could be captured. It is also efficient in handling long-term dependencies and has the potential to outperform others in terms of speed and capturing complex patterns.

3.6. **Evaluation of metrics performance.** To compare performance between different models, including error and regression metrics analysis for predicting the Dst index. Considering the lowest values of error along with the highest correlation in the measured metrics of the Dst index suggests that the model has better performance in estimating future Dst index values. Three indicators are used to evaluate the forecasting performances using MSE, MAE, RMSE, MAPE, sMAPE and MASE by comparing the actual Dst and forecasted Dst. Taking y_{act} as the actual Dst at time t, y_{fct} as the forecasting value at time t, and N as the number of data samples, the metrics will be calculated as the following:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(y_{act} - y_{fct})^2 \tag{14}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y_{act} - y_{fct}| \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_{act} - y_{fct})^2} \tag{16}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_{act} - y_{fct}|}{y_{act}} \tag{17}$$

$$sMAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_{act} - y_{fct}|}{(|y_{act}| + |y_{fct}|)/2} \tag{18}$$

$$MASE = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_{act} - y_{fct}|}{\frac{1}{N-1}\sum_{i=2}^{N}|y_{act} - y_{fct} - 1|} \tag{19}$$

4. Results and Discussion.

4.1. **Data description and platform development.** The available Dst index data are retrieved from the International Service of Geomagnetic Indices (ISGI)[1] with a joint institute collaboration of the World Data Center for Geomagnetism, Kyoto[2]. Geomagnetic data analysis contains 327264 hours of data from January 9, 1986, to December 31, 2023, covering four solar cycles (SC22-SC25) or equivalent to 37 years. These data samples are partitioned into 90%-10% as training and testing sets simulated in the Python environment. During the training set, a cross-validation technique via rolling window analysis, available as the TimeSeriesSplits class in the scikit-learn library, was incorporated to train model development. The model hyperparameters were tuned, and then finally, the test set with the unseen test set compared the model performances. Therefore, preserving the temporal data sequence ensures robust and generalized model development.

4.2. **Ablation analysis of multistep forecasting results.** Examining the performance of the Dst index is hourly forecasted from the horizon of 1-, 3-, 6-, 9-, and 12-steps ahead of performances of the predicted Dst index. Comparisons among different step-ahead performances were conducted to observe how the error metric performances varied (see Figure 5). In this work, the best tuned hyperparameters were obtained based on these configurations: number of hidden units per layer of 256, learning rate of 0.01, batch size of 32, windows size of 24, number of epochs of 100, Adam optimizer and L2 regularization.

The superior, preferable forecasted Dst index can be selected through the lowest error metrics which provide a comprehensive measure of accuracy to the produced forecasts. Apparent trends could be observed, such as all error metrics decreasing at shorter horizons. It is noteworthy that the N-BEATS model consistently exhibits the least variability.

[1]https://isgi.unistra.fr/index.php
[2]https://wdc.kugi.kyoto-u.ac.jp/

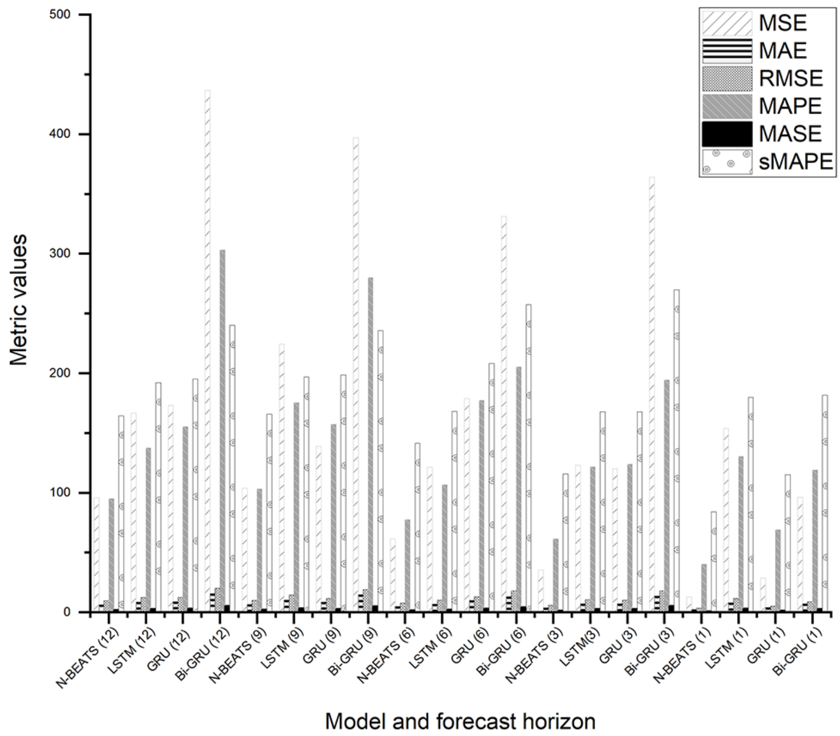


FIGURE 5. Performance metrics to evaluate the model structure with different forecast horizons

The recorded values for MSE, MAE, RMSE, MAPE, sMAPE, and MASE, respectively, were 12.92, 2.36, 3.58, 40.21, 84.07, and 0.97 during 1-step ahead. However, the forecast horizon kept increasing into a longer duration of 12-step ahead, showing performances were also among the lowest compared to other models with the least MSE, MAE, RMSE, MAPE, sMAPE, and MASE, respectively, of 96.01, 6.23, 9.76, 94.98, 164.47 and 2.13. N-BEATS steadily outperforms other models in terms of the most stable performance across different horizons, demonstrating minimal error metrics. LSTM and GRU demonstrated comparable error metrics performance, with GRU sometimes outperforming LSTM, which fluctuates inconsistently as the forecast horizon increases. This given by GRU surpassed LSTM for 1-step ahead forecasts across all metrics, which was more effective in short-term forecasting. GRU showed MSE, MAE, RMSE, MAPE, sMAPE, and MASE, respectively, of 28.64, 3.89, 5.11, 69.08, 115.28, and 1.56. Meanwhile, LSTM showed MSE, MAE, RMSE, MAPE, sMAPE, and MASE, respectively, of 153.83, 7.93, 11.73, 130.33, 179.97 and 3.30. Moreover, LSTM dominates to perform better than GRU for 6 to 12-step ahead forecast horizons, showing LSTM capabilities at capturing patterns of long-term dependencies. The Bi-GRU demonstrated a significant trend of underperforming poorly the most when compared to other models. It even demonstrated the highest variability among other models between 1-step (i.e., MSE, MAE, RMSE, MAPE, sMAPE, and MASE, respectively, of 96.15, 7.02, 8.87, 119.20, 181.56, 2.82) and 12-step (i.e., MSE, MAE, RMSE, MAPE, sMAPE and MASE, respectively, of 436.59, 15.72, 20.15, 303.35, 240.12 and 5.42) ahead forecast horizon. Hence, N-BEATS demonstrated superior performance with the lowest error metrics across all forecast horizons among all tested models.

Computational processing time is also presented in Figure 6. Analysis of training and testing time was recorded to examine the computing performance required when executing the deep learning models. It can be observed that all models except N-BEATS spend thousands of seconds training the deep learning model for Dst index forecasting. Training time indicates that stable training time increases gradually as the forecast horizon expands. Thus, the performance results are much more consistent. Huge computation resources

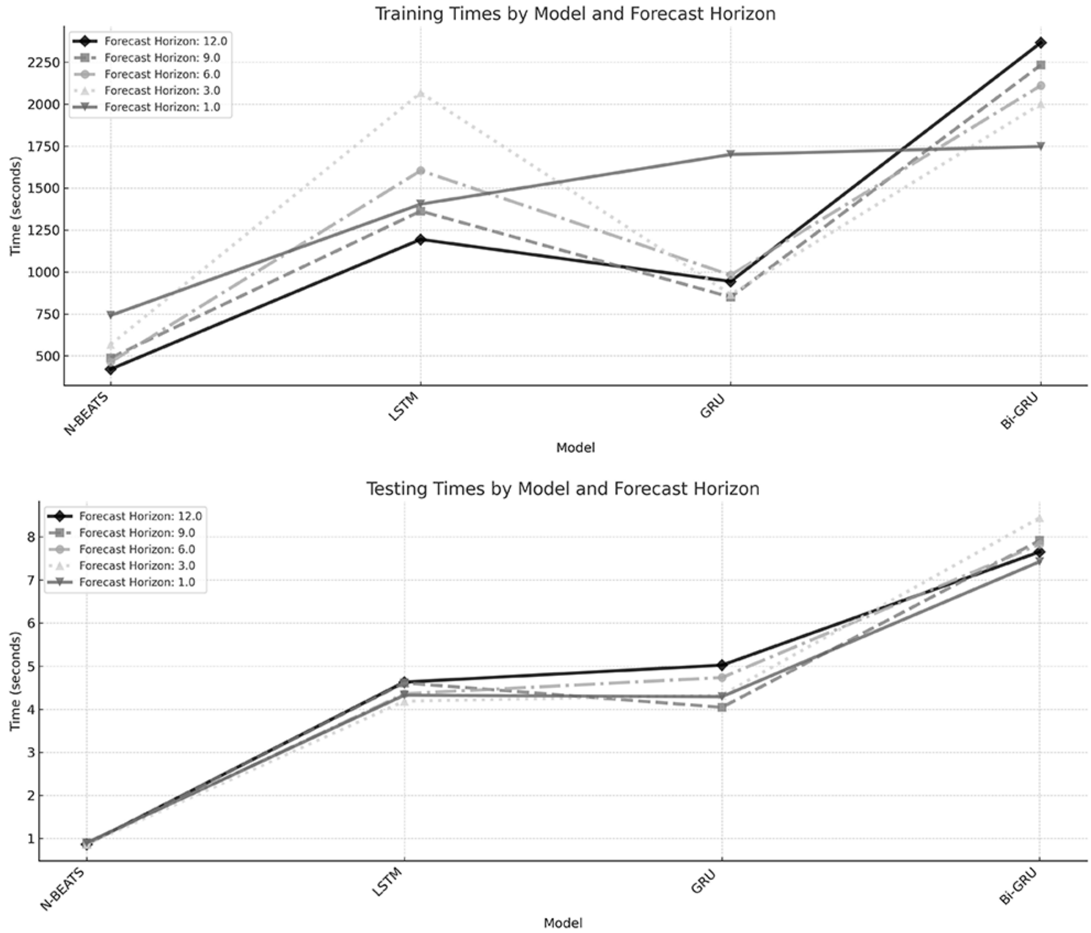
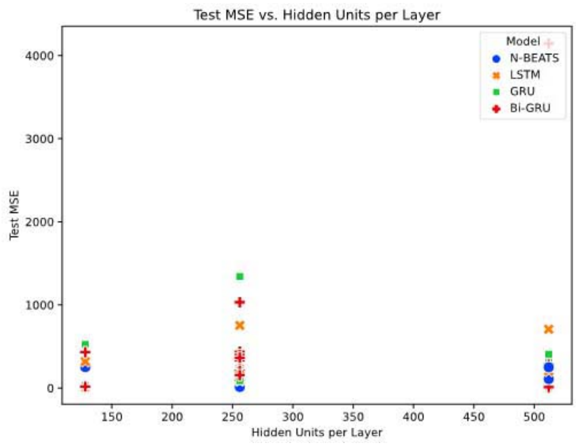


FIGURE 6. Evaluation of computation processing time

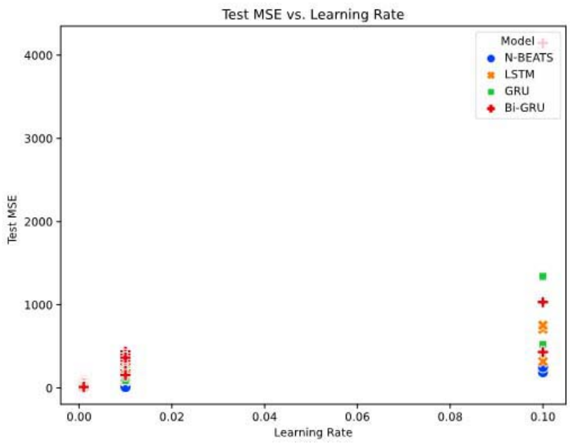
with an average time of 2091.76 seconds by Bi-GRU have significantly peaked amongst all models, indicating high computational cost for longer horizons due to the model's complexity in processing bidirectional data in both forward and backward directions. For the testing time, a similar pattern was observed, such that Bi-GRU consumes the largest computational testing time, N-BEATS used the least and most consistent testing time, while LSTM and GRU have moderate testing times.

Figure 7 depicts the hyperparameters setting influences on the error of MSE. It denotes that Bi-GRU and GRU were the most sensitive to the change in the learning rate. Most of the models respond to low MSE, i.e., 0.01 when the learning rate was implemented in the model, while vice versa, as the learning rate increased to 0.10, there was a significant increase in the MSE. Similar to the learning rate, both Bi-GRU and GRU are also sensitive to variation of MSE, which could be observed when changing window size. More stable MSE variation could be observed for N-BEATS to remain low regardless of the setting for hyperparameter changes. Notably, both LSTM and GRU performed at intermediate levels; GRU was prone to more sensitive hyperparameter changes in all setting conditions with slightly higher values than LSTM.

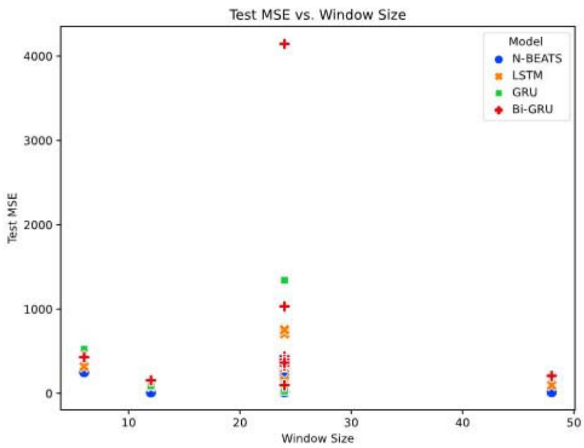
4.3. **Geomagnetic storm events analysis.** To visualize the forecasted Dst compared to the actual Dst, a comparison between these predicted and the observed values plotting and its regression metrics, namely correlation coefficient (CC) and multiple determination coefficient (R^2), was examined, as shown in Figure 8. Examining the performances throughout the test data on the selected geomagnetic storms on November 4, 2021 [21], February 4, 2022 [22], and April 24, 2023 [23], which respectively present moderate, minor, and severe storms, was tested. Table 1 summarizes CC and R^2, which were examined



(a)



(b)



(c)

FIGURE 7. MSE and hyperparameter sensitivity analysis of (a) the number of hidden units, (b) learning rate, and (c) window size

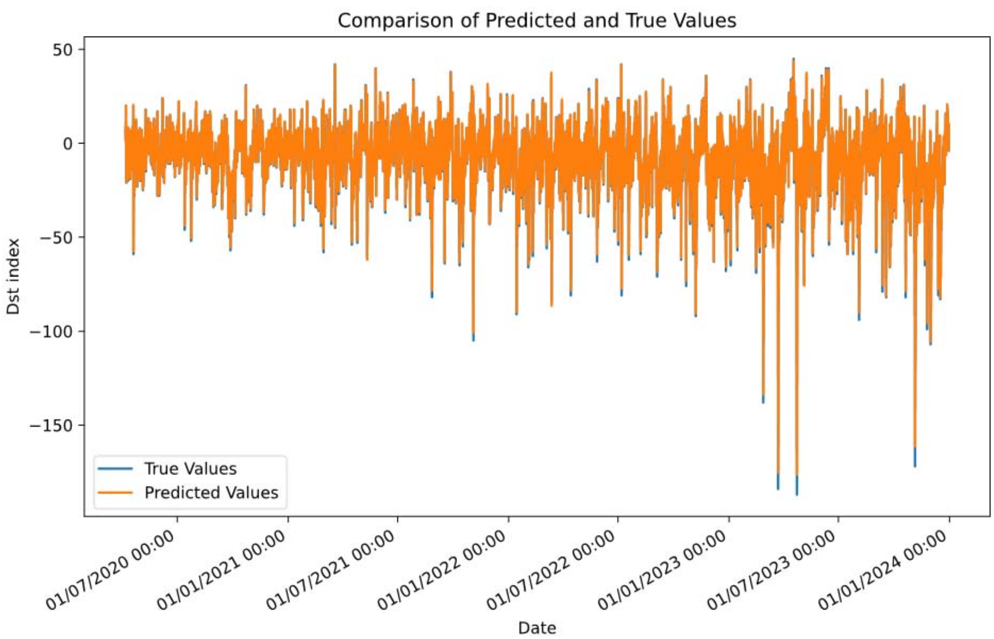


FIGURE 8. Comparison of the dynamics of Dst plotting to compare between actual Dst and forecasted Dst for the test dataset

Table 1. Comparison of the N-BEATS performance of the Dst index from April 7, 2020, to December 31, 2023, with the predictions that were made 1, 3, 6, 9, and 12 steps ahead. A comparison of selected geomagnetic storm events in solar cycle 25 was also included.

	Overall test data	Nov 4th, 2021	Feb 4th, 2022	Apr 24th, 2023
Correlation coefficient (horizon)				
CC (1)	0.9711	0.9648	0.9664	0.9893
CC (3)	0.8736	0.8588	0.8383	0.9305
CC (6)	0.7623	0.7019	0.6424	0.8100
CC (9)	0.6760	0.5440	0.4058	0.6917
CC (12)	0.5995	0.4102	0.2541	0.5221
Multiple determination coefficient (horizon)				
R^2 (1)	0.9431	0.9308	0.9340	0.9786
R^2 (3)	0.7631	0.7376	0.7028	0.8658
R^2 (6)	0.5811	0.4927	0.4127	0.6561
R^2 (9)	0.4570	0.2959	0.1647	0.4785
R^2 (12)	0.3593	0.1683	0.0645	0.2725

across the range of horizon between 1 and 12 steps ahead of forecasting. These indicators are essential to demonstrate the prediction strength and direction between actual and forecasted Dst index with higher values, which is normally preferable in model selection. It demonstrates the Dst index forecasting of a step ahead as the best-performed model with N-BEATS. The highest performance of CC = 0.9711 and $R^2 = 0.9431$ demonstrates a strong predictive ability at the 1-step ahead of horizon. This correlation is observed to be almost consistent across geomagnetic storms over time. As the horizon increases, these metrics exhibit a decreasing pattern with increasing horizons, implying weaker predictive power. However, it was notable that at least 70% and 50% were presented in both CC and R^2 performances for at least a 9-step ahead of horizon.

5. **Conclusions.** To conclude, the proposed novel model of Dst index forecasting using N-BEATS outperformed other models with high accuracy and improved the forecasting in comparison. The N-BEATS model is capable of accurately tracking the geomagnetic storm events for up to 97% correlation, 95% multiple determination coefficient, and lowest error metrics. This includes MSE, MAE, RMSE, MAPE, sMAPE, and MASE, among all deep network models. Moreover, the N-BEATS model has proven to add advantages in terms of fast computing training and testing time, minimal error variability with good consistencies, and generalization properties to learn the patterns, which have performed remarkably well. Future research will extend the Dst index prediction onto long-term forecasting, providing a more accurate lead to better prediction outcomes. In addition, implementing advanced predictions, which require limited computational resources in real-time operational forecasting, will benefit from the faster prediction process, which may be considered.

Acknowledgment. This research is part of the Fundamental Research Grant Scheme, FRGS (ref: FRGS/1/2024/TK08/UPM/02/2) approved by the Ministry of Higher Education Malaysia and is currently supported by Universiti Teknologi MARA under the GPM grant (ref: 600-RMC 5/3/GPM (047/2023)). We would like to acknowledge the national institutes involved, the INTERMAGNET network, and ISGI (isgi.unistra.fr).

REFERENCES

[1] T. Alberti, D. Faranda, G. Consolini, P. De Michelis, R. V. Donner and V. Carbone, Concurrent effects between geomagnetic storms and magnetospheric substorms, *Universe*, vol.8, no.4, 226, 2022.

[2] I. A. Oludehinwa, O. I. Olusola, O. S. Bolaji, O. O. Odeyemi and A. N. Njah, Magnetospheric chaos and dynamical complexity response during storm time disturbance, *Nonlinear Process. Geophys.*, vol.28, no.2, pp.257-270, 2021.

[3] A. Lethy, M. A. El-Eraki, A. Samy and H. A. Deebes, Prediction of the Dst index and analysis of its dependence on solar wind parameters using neural network, *Sp. Weather*, vol.16, no.9, pp.1277-1290, 2018.

[4] F. Nuraeni, M. Ruhimat, M. A. Aris, E. A. Ratnasari and C. Purnomo, Development of 24 hours Dst index prediction from solar wind data and IMF Bz using NARX, *J. Phys. Conf. Ser.*, vol.2214, no.1, 12024, 2022.

[5] I. N. Myagkova, V. Shiroky and S. Dolenko, Prediction of geomagnetic indexes with the help of artificial neural networks, *E3s Web Conf.*, vol.20, 2011, 2017.

[6] A. Gruet, M. Chandorkar, A. Sicard and E. Camporeale, Multiple-hour-ahead forecast of the Dst index using a combination of long short-term memory neural network and Gaussian process, *Sp. Weather*, vol.16, no.11, pp.1882-1896, 2018.

[7] P. Wintoft and M. Wik, Exploring three recurrent neural network architectures for geomagnetic predictions, *Front. Astron. Sp. Sci.*, vol.8, 664483, 2021.

[8] Wihayati, H. D. Purnomo and S. Trihandaru, Disturbance storm time index prediction using long short-term memory machine learning, *Proc. of the 4th International Conference on Computer and Informatics Engineering (IC2IE)*, pp.311-316, 2021.

[9] A. Hu, E. Camporeale and B. Swiger, Multi-hour-ahead Dst index prediction using multi-fidelity boosted neural networks, *Sp. Weather*, vol.21, no.4, 2023.

[10] I. Gulati, H. Li, M. Johnston and S. Dlay, Classification based detection of geomagnetic storms using LSTM neural network, *2022 3rd URSI Atlantic and Asia Pacific Radio Science Meeting (AT-AP-RASC 2022)*, pp.1-4, 2022.

[11] M. J. Alam, J. Rabbi and S. Ahamed, A deep neural network based model to forecast geomagnetic storm and observe the effects of various activation functions on the model, *2022 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET)*, pp.1-5, 2022.

[12] J. Zhang, Y. Feng, J. Zhang and Y. Li, The short time prediction of the Dst index based on the long-short time memory and empirical mode decomposition-long-short time memory models, *Appl. Sci.*, vol.13, no.21, 11824, 2023.

[13] A. O. Efitorov, I. N. Myagkova, V. R. Shirokii and S. A. Dolenko, The prediction of the Dst-index based on machine learning methods, *Cosm. Res.*, vol.56, no.6, pp.434-441, 2018.

[14] A. Tasistro-Hart, A. Grayver and A. Kuvshinov, Probabilistic geomagnetic storm forecasting via deep learning, *J. Geophys. Res. Sp. Phys.*, vol.126, no.1, 2021.

[15] Y. Abduallah, J. T. L. Wang, P. Bose, G. Zhang, F. Gerges and H. Wang, Forecasting the disturbance storm time index with Bayesian deep learning, *Proceedings of the International Florida Artificial Intelligence Research Society Conference*, Florida, OJ, United States, vol.35, 2022.

[16] S. B. Xu, S. Y. Huang, Z. G. Yuan, X. H. Deng and K. Jiang, Prediction of the Dst index with bagging ensemble-learning algorithm, *The Astrophysical Journal Supplement Series*, vol.248, no.1, 14, 2020.

[17] M. Cristoforetti, R. Battiston, A. Gobbi, R. Iuppa and M. Piersanti, Prominence of the training data preparation in geomagnetic storm prediction using deep neural networks, *Sci. Rep.*, vol.12, no.1, 2022.

[18] Y. Y. Li et al., Selection of the main control parameters for the Dst index prediction model based on a layer-wise relevance propagation method, *The Astrophysical Journal Supplement Series*, vol.260, no.1, 6, 2022.

[19] J. A. Lazzús, P. Vega, P. Rojas and I. Salfate, Forecasting the Dst index using a swarm-optimized neural network, *Sp. Weather*, vol.15, no.8, pp.1068-1089, 2017.

[20] B. Oreshkin, G. Dudek, P. Pełka and E. Turkina, N-BEATS neural network for mid-term electricity load forecasting, *Appl. Energy*, vol.293, 116918, 2021.

[21] T. M. Matamba, D. W. Danskin, R. R. Nndanganeni and M. Tshisaphungo, Space weather impacts on the ionosphere over the southern African mid-latitude region, *Earth, Planets Sp.*, vol.75, no.1, 142, 2023.

[22] R. Bojilova and P. Mukhtarov, Comparative analysis of global and regional ionospheric responses during two geomagnetic storms on 3 and 4 February 2022, *Remote Sens.*, vol.15, 1739, 2023.

[23] S. S. K. Rajana et al., Impact of two severe geomagnetic storms on the ionosphere over Indian longitude sector during March-April 2023, *Astrophys. Space Sci.*, vol.369, no.1, 3, 2024.