JSST
Journal of Space Science and Technology

P-ISSN: 2008-4560
E-ISSN: 2423-4516

Journal of Space Science and Technology

Vol. ??, No. ?, pp. 1-16, 2026
https://doi.org/10.22034/jsst.2025.1585
Journal Homepage: https://jsstpub.com

Original Research Paper

Machine Learning–Driven Prediction of the S4 Scintillation Index and Geomagnetic Storms using GNSS-RO Data

Delaram Fattahi[1], Masoud Khoshsima[2], Saber Karami[3], Hadi Jalili[4], Jamal Aghayari[5], and Sahar Barzegar [6,*]

1,2. Institute of Geophysics, University of Tehran, Tehran, Iran
3-6. Iranian Space Research Center, Tehran, Iran

ARTICLE INFO

Article History:

Received dd mmm yyyy
Revisee dd mmm yyyy
Accepted dd mmm yyyy
Available Online dd mmm yyyy

Keywords:

Machine learning
Random forest
Scintillation index
GNSS radio occultation
Geomagnetic storm

ABSTRACT

This study applies the Random Forest (RF) machine learning (ML) algorithm to GNSS radio occultation (GNSS-RO) data for two critical space weather forecasting tasks: the prediction of ionospheric amplitude scintillation (S4 index) and the classification of geomagnetic storm occurrence. For S4 index prediction, an RF regression model was trained on a comprehensive feature set derived from interplanetary magnetic field components, geomagnetic indices, solar radio flux, and historical S4 statistics. The model's architecture is optimized, and Recursive Feature Elimination with Gini Importance (Mean Decrease Impurity) method is applied to identify the most predictive features. For storm detection, an RF classifier was trained on parameters including Total Electron Content (TEC), S4 indices, and key solar and geomagnetic variables to distinguish between storm and non-storm conditions. The optimized S4 prediction model achieved high precision with a Mean Absolute Error of 0.007263 and an R^2 score of 0.517352. Feature selection via Gini Importance significantly improved model efficiency, increasing the Adjusted R^2 by 50.5%. The geomagnetic storm classifier demonstrated a critical strength of recall of 0.99 for storm events, ensuring missing only 1% of actual storms. Analysis of feature importance confirmed the model's physical validity, correctly identifying the Kp index, F10.7 solar flux and Dst index as the primary drivers for prediction, which aligns with established solar-terrestrial physics. The results demonstrate the high potential of machine learning, specifically Random Forests, for precise space weather forecasting.

* Corresponding Author's E-mail: sbarzegar@ut.ac.ir

How to Cite this Article:
D. Fattahi, M. Khoshsima, S. Karami, H. Jalili, J. Aghayari, and S. Barzegar, "Machine learning–driven prediction of the S4 scintillation index and geomagnetic storms using GNSS-RO data," *Journal of Space Science and Technology*, Vol. ??, No. ?, pp. 1-16, 2026, https://doi.org/10.22034/jsst.2025.1585.

1 INTRODUCTION

Geomagnetic storms, originating from complex solar wind-magnetosphere interactions, pose a significant threat to critical technological systems such as communications, navigation, and electrical infrastructure [1, 2]. Therefore, the accurate prediction of geomagnetic storms is essential for developing effective mitigation strategies and ensuring operational preparedness.

One of the most important parameters that helps assessing geomagnetic storms impacts is the S4 ionospheric scintillation index. Ionospheric scintillation denotes rapid fluctuations in the amplitude and phase of radio waves caused by electron density irregularities in the ionosphere. These irregularities, often driven by geomagnetic storms, alter the medium's refractive index, disrupting signal propagation [3-5]. This phenomenon is of significant scientific and practical concern, as it adversely impacts critical space-based systems, including satellite navigation, military and civilian telecommunications, and launch vehicle operations [6-14]. Consequently, developing reliable predictive models for scintillation has become increasingly imperative [15]. The S4 index, a dimensionless measure of scintillation intensity, is defined as the normalized standard deviation of the signal intensity, as given by [16, 17]:

$$S4 = \sqrt{\frac{\langle I^2 \rangle - \langle I \rangle^2}{\langle I \rangle^2}}. \quad (1)$$

While valuable, traditional statistical and physics-based models often fail to capture the non-linear dynamics of ionospheric scintillation. This limitation has motivated growing interest in ML approaches.

Machine learning has recently emerged as a powerful tool for analyzing and predicting complex systems, demonstrating significant potential for generating accurate forecasts by identifying complex patterns within large datasets [18, 19]. Among these techniques, the RF algorithm has become a dominant technique as an effective ensemble method for both classification and regression tasks [20]. The inherent ensemble architecture of the RF algorithm presentss robustness, allowing it to model complex non-linear relationships and avoid overfitting. Consequently, RF can extract important features from data, which supports the generation of accurate and generalizable predictions. Other advanced ML approaches such as Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) have also been applied; however, previous studies indicate that they often entail high computational demands and remain prone to overfitting [21–26].

In this study, we apply a random forest model to global GNSS-RO observations to forecast both the S4 scintillation index and the onset of geomagnetic storms. We assemble a dataset by pairing GNSS-RO soundings from COSMIC-2 satellites and key solar-geomagnetic indices (Dst, Kp, F10.7, IMF components). The data were selected to encompass seven geomagnetic storm periods between 2023 and 2025, covering their active days from Day of Year 074, 2023, to Day of Year 002, 2025 (as listed in Table 1). Our results demonstrate that a Random Forest algorithm can effectively predict the S4 scintillation index, with performance enhanced through feature selection. A dedicated storm-prediction model was trained to classify both the presence and absence of geomagnetic storms, achieving recall of 0.99 for storm events, ensuring missing only 1% of actual storms. Feature importance analysis via Gini Importance highlighted the Dst index, Kp index, and temporal parameters as critical drivers, optimizing the model to an 50.5% increase in Adjusted R^2. Our findings demonstrate the strong synergy between GNSS-RO data and ML for reliable space weather forecasting and risk mitigation for communication and navigation infrastructure.

2 Related Works

In recent years, the use of ML for the prediction and analysis of ionospheric disturbances has become a thriving field of research. Early studies, such as the work by Das et al. (2010), employed simple neural networks for climatological and long-term forecasts. While they could reproduce seasonal scintillation patterns, they were unable to capture dynamic day-to-day fluctuations [27]. Subsequent research shifted towards more robust algorithms and ensemble learning methods. For example, Rezende et al. (2010) provided improved hourly predictions of the S4 index using a Bagging-CART model [28], while Nasurudiin et al. (2024) achieved high accuracy in S4 prediction by comparing RF and

XGBoost and introducing a new physical precursor (Pc5 waves), thereby highlighting the importance of feature engineering [29].

One of the most comprehensive studies in this domain is the work of Carvalho et al. (2022). By constructing a rich database of 27 features from various sources and applying a rigorous methodology, they achieved exceptionally high accuracy (R^2=0.978) in 30-minute S4 predictions using RF [30]. These studies demonstrate the power of combining multi-modal data with modern ensemble algorithms. Concurrently, some researchers reformulated the problem from regression to classification; for instance, Darya et al. (2022) successfully classified scintillation intensity levels using bagged trees [31].

With the advent of deep learning, the frontiers of this field have shifted, with a growing focus on spatio-temporal modeling. Innovative research, such as the work by Atabati et al. (2024) with their ConvGRU architecture [32] and Tian et al. (2022, 2024) using a 3D U-Net architecture [33, 34], aims to generate two-dimensional maps of the S4 index rather than single-point predictions. By simultaneously learning spatial and temporal patterns, using past S4 values (autoregressive models), or extracting latent relationships between the lower atmosphere and the ionosphere, these models have achieved unprecedented results in understanding and predicting the global distribution of irregularities.

The present study is situated within this scientific landscape, with both commonalities and distinctions. Our classification model, like modern studies, utilizes the RF algorithm. However, it introduces a key difference: instead of predicting the S4 index itself (the effect), it directly aims to predict the cause of its variation the geomagnetic storm (the cause). In this approach, ionospheric data serves as one of several physical precursors. On the other hand, our regression model aims to predict the S4 index, understand the contemporaneous relationships between variables, and methodologically evaluate the Gini Importance feature selection technique. By employing standard techniques while offering a unique perspective, both of our models contribute to the ongoing research in this field.

3 Data and Methodology

This section details the acquisition process for the data used in this paper. Additionally, the RF ML algorithm and its evaluation methods, which were employed to predict both geomagnetic storms and the S4 index, are explained in detail.

3.1 Data Acquisition

In this study, Radio Occultation (RO) data were extracted from the COSMIC-2 satellite constellation (https://www.cosmic.ucar.edu/what-we-do/cosmic-2/data). The data were selected to encompass seven geomagnetic storm periods between 2023 and 2025, covering their active days from Day of Year 074, 2023, to Day of Year 002, 2025 (as listed in Table 1). The COSMIC-2 constellation consists of six satellites orbiting at an altitude of 550 km with an inclination of 24 degrees. Furthermore, solar and geomagnetic indices were obtained from NASA's OMNIWeb database (https://omniweb.gsfc.nasa.gov/form/dx4.html).

Table 1: Summary of the expanded dataset including seven geomagnetic storms.

Year	Day of Year	Storm Active Days
2023	074-083	082, 083
2023	105-114	113, 114
2023	301-310	309, 310
2024	124-133	131, 132, 133
2024	217-226	225, 226
2024	276-285	282, 283, 284, 285
2025	359-002*	001, 002

*Note: The data chunk for Storm 7 spans from day 359 of 2024 to day 002 of 2025.

3.2 The RF Algorithm

The RF algorithm is an ensemble learning method that increases predictive accuracy and reduces the risk of overfitting by combining multiple decision trees. The algorithm is designed for two primary ML tasks: regression and classification.

In a classification problem, the objective is to predict a discrete or categorical target variable, such as diagnosing a disease or filtering emails. The success of RF heavily relies on the use of uncorrelated decision trees. If the trees are highly correlated, the final result will be similar to that of a single decision tree. Therefore, the process of generating uncorrelated trees in RF is achieved through two key techniques: bagging (bootstrap aggregating) and feature randomness. In bagging, each decision tree provides a parallel estimate of the output. Simply put, RF builds multiple decision

trees and merges them. This parallel execution of trees also helps to reduce computation time. With feature randomness, a random subset of features is selected for consideration at each split within every decision tree. Ultimately, when making a prediction, each tree casts a vote for the target class, and the class with the majority of votes is chosen as the final result. The criteria for splitting nodes are typically the Gini impurity or entropy, which measure node impurity [35]. For example, this method could be used to predict a patient's probability of developing diabetes based on their age, weight, and blood sugar levels.

In a regression problem, the goal is to predict a continuous variable, such as the price of a house or the temperature. Here, each tree is a regression model, and the final prediction is derived from the average of the predictions from all trees, rather than by voting. The splitting criteria for regression trees are usually the Mean Squared Error (MSE) or the Mean Absolute Error (MAE), which calculate the deviation of predictions from the actual values. As with classification, bagging and random feature selection are used to create diversity among the trees. For instance, a RF regressor can predict housing prices based on factors like square footage, number of rooms, and geographic location.

The main differences between the two types of RFs lie in their output type, evaluation metrics, and tree structure. In classification, the output is a discrete class (e.g., 0 or 1), and metrics such as Accuracy or F1-Score are used for evaluation. In regression, the output is a continuous value, and the evaluation metrics are typically MSE or R^2. Furthermore, classification trees use discrete rules (e.g., yes/no), while regression trees store numerical values in their leaf nodes. Generally, the algorithm follows the steps below to make predictions (Figure 1): Step 1: Select random bootstrap samples from a given training dataset. Step 2: Construct a decision tree for each sample, considering only a random subset of features for each split. Step 3: Obtain a prediction from each tree. In classification, the class predicted by the majority of trees is chosen (voting). In regression, the average of the predictions from all trees is taken. Step 4: The result with the most votes (for classification) or the average value (for regression) is selected as the final prediction. Step 5: The model's performance is evaluated using various metrics appropriate for the task.

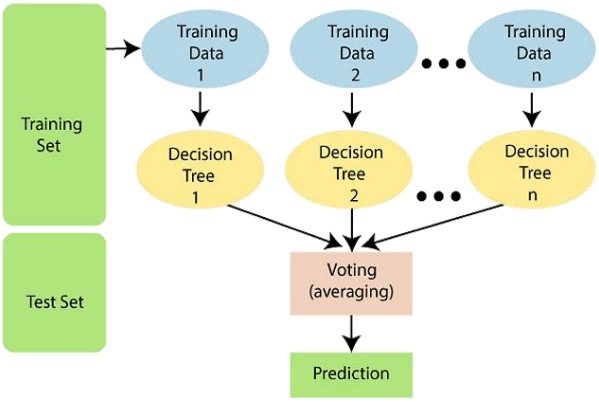

Figure 1. Flowchart of the RF algorithm's steps for data prediction. (Source: RF algorithm in ML. (2023, December 15). AlmaBetter. https://www.almabetter.com/bytes/tutorials/data-science/random-forest)

3.3 Feature Engineering and Selection for the RF Model

In addition to the direct use of GNSS-RO (RO) data and solar-geomagnetic indices, a comprehensive list of features for training the ML model can be generated by applying descriptive and distributional statistical processes to these initial datasets (a process known as feature engineering). For example, four of the statistical operations that can be applied to the primary data to create new features are as follows:
Standard Deviation (σ): Measures the dispersion of a dataset relative to its mean. It is calculated as the square root of the variance using the formula below.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \quad (2)$$

Mean (μ): Represents the central tendency or average of the data, calculated by dividing the sum of the values by the count of the values in a data sample.

$$\bar{X} = \frac{\sum X}{N} \quad (3)$$

Kurtosis: Measures the "peakedness" or "flatness" of a data distribution compared to a normal distribution. It is a statistical metric that indicates how concentrated the data are around the mean and how heavy the tails (outliers) of the distribution are. Pearson utilized the fourth moment to calculate the kurtosis coefficient. If μ is the mean and σ is the standard deviation of a

random variable X, then the kurtosis, Kur(X), is defined as:

$$Kur(X)=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^2} \quad (4)$$

Skewness: Measures the asymmetry of the data distribution around its mean. It is a statistical metric indicating the degree of asymmetry of a probability distribution relative to a symmetric distribution (like the normal distribution). Skewness shows how the data in a set are scattered around their mean and whether they are skewed to one side. It is calculated as follows:

$$Skewness=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^{3/2}} \quad (5)$$

Tables 2 and 5 provide a complete list of the features used to train the RF models for the prediction of the S4 scintillation index and the prediction of storms, respectively.

In this research, the feature selection process was conducted using the RFECV method, based on the RF algorithm. Initially, a primary set of features was extracted from the raw GNSS-RO data and solar-geomagnetic indices using statistical methods (such as mean, standard deviation, skewness, and kurtosis over various time windows) and mathematical transformations (like sine/cosine components of the hour of the day). Subsequently, by employing Gini Importance method, unimportant features were iteratively eliminated to identify an optimal set of 20 key features, including the Dst index, Kp index, components of the interplanetary magnetic field, and temporal parameters. This process led to an 8.94% improvement in the Adjusted R-squared (Adjusted R^2) for S4 prediction and increased the storm classification accuracy to 97%. This demonstrates a reduction in noise and an enhanced focus of the model on the most relevant predictors [36, 37].

3.4 Evaluating the RF

3.4.1 Evaluating the RF Regression Model

This section discusses five well-known metrics for evaluating the results of a regression model, such as the RF algorithm. In the formulas, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}_i$ is the mean of the sample values.

Mean Absolute Error (MAE): This metric measures the average of the absolute differences between the actual and predicted values. A lower value indicates a better model. It treats all errors equally, with each prediction error contributing proportionally to the overall average. This can be useful when all errors are considered equally important. MAE is also robust to outliers and provides an easily interpretable measure of prediction error.

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \quad (6)$$

Mean Squared Error (MSE): This metric measures the average of the squared differences between the actual and predicted values. A lower value indicates a better model. Calculating MSE serves two primary purposes: eliminating negative signs and emphasizing larger errors or outliers by squaring them, which penalizes them more heavily. The sum of the squared differences is divided by the total number of data points (n), ensuring that MSE is comparable across datasets of different sizes. However, because it involves squaring the errors, the units of MSE are not on the same scale as the dependent variable, which makes it less interpretable for communicating the average prediction error in practical terms.

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \quad (7)$$

Root Mean Squared Error (RMSE): This metric is the square root of the Mean Squared Error (MSE). It returns the error metric to the original units of the target variable, making it more interpretable than MSE.

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \quad (8)$$

R-squared (R2) Score: This metric indicates the proportion of the variance in the dependent variable that is explained by the model. Its value ranges between 0 and 1; the closer it is to 1, the better the model.

$$R^2=1-\frac{\sum_i(y_i-\hat{y}_i)^2}{\sum_i(y_i-\bar{y}_i)^2} \quad (9)$$

Adjusted R-squared (Adjusted R2) Score: This metric accounts for the number of predictors (features) in a regression model. It penalizes the excessive use of irrelevant features and can help detect overfitting. It is calculated according to formula (10), where n is the number of test samples and p is the number of predictors.

Adjusted R^2=1- $\frac{(1-R^2)(n-1)}{n-p-1}$ (10)

3.4.2 Evaluating the RF Classification Model

Five common metrics for evaluating the results of a classification model are introduced below [42]. Accuracy: This is the ratio of correctly predicted instances (True Positives, TP, and True Negatives, TN) to the total number of predictions made.

$$Accuracy = \frac{TP+TN}{FP+FN+TP+TN} \quad (11)$$

Precision: This measures the ratio of correct positive predictions (TP) among all positive predictions made by the model. It is calculated as the ratio of True Positives (TP) to the sum of True Positives (TP) and False Positives (FP). It indicates how many of the predicted positive cases were correctly identified. In essence, this metric represents the proportion of correct positive predictions to the total positive predictions (e.g., what fraction of the storm alerts were actually storms?).

$$Precision = \frac{TP}{FP+TP} \quad (12)$$

Recall: This measures the ratio of correct positive predictions (TP) among all actual positive samples in the dataset. It is calculated as the ratio of True Positives (TP) to the sum of True Positives (TP) and False Negatives (FN). It indicates how many of the actual positive cases the model can correctly identify. In essence, this metric represents the proportion of correct positive predictions to the total actual positive cases (e.g., what fraction of the actual storms were identified by the model?).

$$Recall = \frac{TP}{FN+TP} \quad (13)$$

F1-Score: This is the harmonic mean of Precision and Recall. It provides a single metric that combines both, creating a balance between them. This score takes both False Positives (FP) and False Negatives (FN) into account, making it an excellent combined metric for evaluating a model, especially in cases of imbalanced data.

$$F1=2\frac{Precision \cdot Recall}{Precision+Recall} \quad (14)$$

Support: This refers to the number of actual occurrences of each class in the dataset. It is simply the count of samples for each class.

3.5 Data Preparation and Splitting

In this study, the performance of the RF model was assessed by partitioning the chronologically ordered data into two subsets: a development set comprising the first 80% of the timeline and a final holdout test set comprising the remaining 20%. This design preserves the natural order of events, which is critical for time-series applications such as ionospheric scintillation, where temporal dependencies must be respected to avoid data leakage. By ensuring that the model is trained only on past data and evaluated strictly on future data, the method eliminates look-ahead bias and provides a fair, realistic assessment of the model's forecasting capability and its ability to generalize without overfitting. The selection of an 80-20 ratio, a common standard in machine learning research, strikes an appropriate balance between the volume of training data (for learning complex patterns) and testing data (for accurate evaluation). This data selection process was executed after performing the necessary preprocessing steps (such as removing outliers and normalization) to ensure that the data distribution in both subsets was as similar as possible to the original population [33].

The choice of a temporal holdout, rather than random shuffling, is fundamental because it ensures the model is evaluated on its true ability to forecast future states from past observations, the central challenge in space weather prediction. Unlike random splits, which risk contaminating training with future information and introducing data leakage, a temporal split preserves the chronological order of events and reflects real-world forecasting conditions. Although this means the test set comes from a single contiguous period, the extensiveness of our dataset guarantees that the 20% holdout segment is both substantial and representative, capturing portions of multiple storm events. This design not only safeguards against look-ahead bias but also addresses concerns about statistical power, providing a robust and reliable foundation for assessing the model's forecasting skill and generalization capacity. Table 2 provides a summary of the data preprocessing, feature engineering, and the overall training and evaluation process.

Table 2. The Data Preprocessing, Feature Engineering, and Model Training/Evaluation Pipeline.

Phase	Actions Taken
Data Cleaning	Removal of outliers and correction of errors or illogical data points.
Feature Engineering	Extraction of statistical features (mean, standard deviation, etc.); addition of sine/cosine temporal components.
Feature Selection	Application of Gini Importance to identify key features and remove less important ones.
Normalization	Standardization (Z-Score) for normally distributed features (Note: This step is not strictly necessary for the RF algorithm).
Data Splitting	Partitioning the data into training (80%) and testing (20%) sets.
Hyperparameter Tuning	Utilizing the Random Search method with cross-validation.
Final Training	Training a final model with optimal hyperparameters on the entire training dataset.
Model Evaluation	Evaluating the final model on the unseen test dataset.

3.6 Hyperparameter Selection with Random Search

To select the optimal hyperparameters, we employed the RandomizedSearchCV technique with time-series cross-validation for our model. Unlike Grid Search, which exhaustively tests all possible combinations of hyperparameters, this method operates more intelligently by randomly sampling a fixed number of combinations (n_iter) from a defined search space. For each sampled combination, the model's performance is evaluated using Time-Series Cross-Validation (TimeSeriesSplit) to ensure that in each fold, the validation data is chronologically after the training data. The optimization metric was the F1-Score, which is more suitable than accuracy for imbalanced datasets (as storms are rare events). By assigning a score to each combination, this process ultimately outputs the set of hyperparameters that achieved the highest performance score. The primary advantage of this method is its significantly higher computational efficiency compared to an exhaustive search, especially when the search space is large.

Therefore to ensure the temporal integrity of our validation process, we employed Time-Series Cross-Validation using the "TimeSeriesSplit" method. This approach explicitly respects the chronological order of the data by creating multiple splits where the training set always precedes the validation set in time. In each successive fold, the training window expands while maintaining a fixed-size validation window that immediately follows the training period. This method effectively prevents data leakage from future to past observations and provides a realistic assessment of the model's performance on unseen temporal data, closely mimicking real-world forecasting scenarios where models predict future events based on historical patterns.

In the RF model, each hyperparameter plays a specific role in controlling the model's structure and behavior. The n_estimators hyperparameter determines the total number of independent decision trees built in the forest. The complexity of each individual tree is managed by three key hyperparameters: max_depth limits the number of levels in the tree, min_samples_split defines the minimum number of samples a node must have to be split, and min_samples_leaf sets the minimum number of samples required in a terminal leaf node. Furthermore, for the RF Classifier, two additional hyperparameters regulate decision-making: criterion sets the function used to measure the quality of a split, while class_weight is used to assign different weights to classes, which is crucial for handling imbalanced datasets.

To identify the optimal hyperparameters, we conducted a search with 50 iterations for both the storm classification model and the S4 regression model, striking a balance between computational efficiency and comprehensive exploration of the parameter space. Model validation was carried out using a five-fold TimeSeriesSplit procedure, in which each validation fold chronologically followed its corresponding training fold. This design preserved the temporal order of the data, thereby preventing data leakage and ensuring that the evaluation reflected realistic forecasting conditions. By structuring the validation in this way, the models were consistently tested on future data relative to their training sets, providing a robust and fair assessment of their predictive performance.

4 Results and Findings

4.1 Prediction of the S4 Amplitude Scintillation Index

This section describes the training and evaluation of an RF model developed to predict the

S4 amplitude scintillation index. The RF algorithm was trained using the features listed in Table 3. We then assess the model's performance in reproducing S4 fluctuations under storm-time conditions.

The optimal hyperparameters selected for the RF regression model indicate a highly complex and granular architecture. The value of n_estimators=210 means the model utilizes 210 individual decision trees to achieve a stable and robust prediction. A max_depth=10 allows each tree to grow quite deep, enabling it to learn very complex and non-linear patterns within the data. The values for min_samples_split=24 (minimum samples to split a node) and min_samples_leaf=19 (minimum samples in a leaf) apply the least possible restrictions, permitting the trees to continue partitioning until they reach the purest possible state.

Table 3. Features Used for S4 Amplitude Scintillation Index Prediction.

Feature	Description
BX_first	First hourly value of the interplanetary magnetic field X-component.
BY_first	First hourly value of the interplanetary magnetic field Y-component.
BZ_first	First hourly value of the interplanetary magnetic field Z-component.
Kp_first	First hourly value of the Kp index.
Dst_first	First hourly value of the Dst index.
F10.7_first	First hourly value of the F10.7 solar radio flux.
BX_mean_3hr	3-hour moving average of the BX component.
BX_min_6hr	3-hour moving average of the BX component.
BX_std_3hr	3-hour moving standard deviation of the BX component.
Dst_max_12hr	12-hour moving maximum of the Dst index.
Dst_skew_6hr	6-hour moving skewness of the Dst index distribution.
Kp_var_6hr	6-hour moving variance of the Kp index.
F10.7_kurt_12hr	12-hour moving kurtosis of the F10.7 flux distribution.
B_total	Total magnetic field magnitude, calculated from BX, BY, and BZ.
sin_minute	Sine component of the minute of the day (to capture diurnal periodicity).
cos_minute	Cosine component of the minute of the day (to capture diurnal periodicity).
S4_std	Standard deviation of the S4 index.
S4_max	Maximum value of the S4 index.
S4_mean	Mean value of the S4 index
TEC_mean	Mean value of Total Electron Content (TEC)
TEC_std	Standard deviation of TEC
TEC_max	Maximum value of TEC
occheight_mean	Mean value of the occultation height
elevation_mean	Mean value of the elevation angle
Proton_Density_first	First hourly value of the Proton Density.
SW_Speed_first	First hourly value of the Solar Wind Speed.
Flow_pressure_first	First hourly value of the Flow Pressure.
R_No_first	First hourly value of the R Number
AE_first	First hourly value of the Auroral Electrojet (AE) index.
S4_mean_lags	S4 mean index lagged by 1, 5, 10, 15, 20, 25, 30 and 60 minute.
BX_mean_2hr	2-hour moving average of the BX component.
Dst_max_3hr	3-hour moving maximum of the Dst index.
imf_clock_angle	Interplanetary Magnetic Field (IMF) Clock Angle (arctan2(B_Y/B_Z)).
B_T	Magnitude of the magnetic field in the Y-Z plane ($\sqrt{(B\_Y^2 + B\_Z^2)}$).
Kp_x_AE	Interaction feature: Product of Kp and AE indices (Kp×AE).
day_of_year	Day of the year (1 to 365/366) derived from datetime_artificial.
sin_day	Sine component of the day of the year (to capture annual periodicity).
cos_day	Cosine component of the day of the year (to capture annual periodicity).

To evaluate the model's performance, the metrics MAE, MSE, RMSE, R^2 Score, and Adjusted R^2 Score, which were previously detailed in Section 3, are used. The results for these five criteria are presented in Table 4 to provide a more precise understanding of the model's effectiveness.

According to the results (shown in Table 4), the MAE value is very small, which generally indicates a low average difference between the predicted and actual values for the test data. The MSE value is also exceptionally low; comparing it to the MAE suggests the presence of very few extreme outliers. On the other hand, the R^2 score

of 0.514 indicates moderate explanatory power, while the minimal difference between the R^2 and Adjusted R^2 scores (0.514377 vs. 0.513379) suggests that the model does not suffer from significant overfitting due to excessive features. Nevertheless, the very low values for MAE, MSE, and RMSE, when considered in the context of the S4 index's scale, indicate a high degree of precision in the predictions. Figure 2 illustrates Feature importance ranking for predicting the mean S4 index. It shows the accuracy of predicting the S4 amplitude scintillation index is most dependent on the lags of the S4 index data itself (S4_mean_lags) and the F10.7 and Bx indices. Therefore, we will remove some of the less impactful features and retrain the model. A new table of evaluation metrics (Table 5) will be generated for the model with fewer inputs to examine the effect of removing these redundant features. The following section explains how more suitable and effective features were selected.

Table 4. Evaluation Metrics for the S4 Prediction Model (Before Feature Selection).

Comparative Parameters	Value
Mean Absolute Error (MAE)	0.007305
Mean Squared Error (MSE)	0.000089
Root Mean Squared Error (RMSE)	0.009422
R-squared (R2) Score	0.514377
Adjusted R-squared (Adjusted R2) Score	0.513379

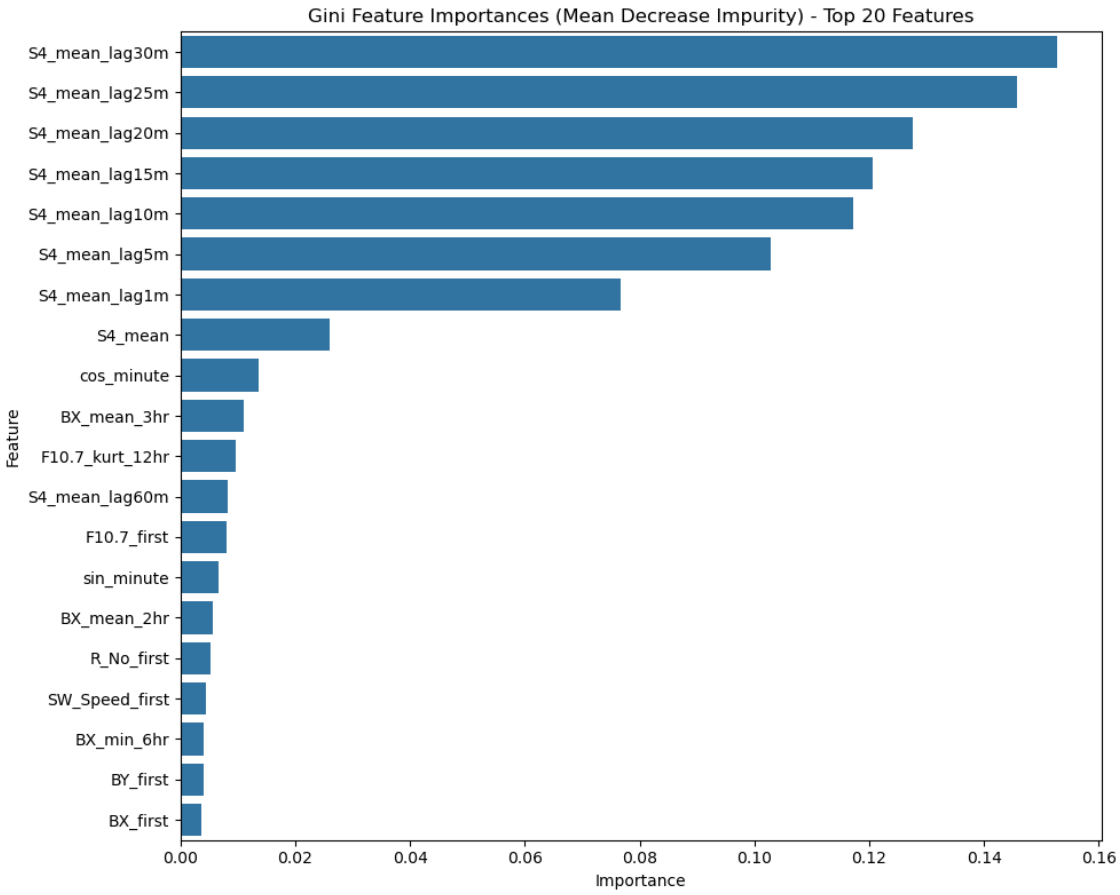

Figure 2. Feature importance ranking for predicting the mean S4 index.

4.1.1 Feature Filtering and Selection of Effective Features

To determine the optimal features, we employed the Gini Importance method, also known as Mean Decrease Impurity (MDI). This approach leverages the intrinsic feature importance metrics generated by Random Forest models, which quantify each feature's contribution to reducing impurity in decision trees. The Gini Importance measures how much each feature decreases the weighted impurity (typically Gini impurity for classification or variance for regression) across all trees in the forest. Features that provide larger impurity reductions when splitting nodes are considered more important, as they better separate target classes or explain variance in regression tasks [38, 39].

The feature selection process involved training a Random Forest model on the complete set of initial features and extracting the Gini Importance scores for each feature. These scores represent the normalized total reduction of impurity brought by each feature across all decision trees in the ensemble. The features were then ranked in descending order based on their importance scores, and the top 20 most influential features demonstrating the highest predictive power for our forecasting model were selected for final model training. This method provides a robust, efficient approach to feature selection that directly leverages the ensemble learning mechanism of Random Forest algorithms.

Table 5. Evaluation Metrics for the S4 Prediction Model (After Selection of Effective Features).

Comparative Parameters	Value
Mean Absolute Error (MAE)	0.007263
Mean Squared Error (MSE)	0.000088
Root Mean Squared Error (RMSE)	0.009393
R-squared (R2) Score	0.517352
Adjusted R-squared (Adjusted R2) Score	0.516857

Ultimately, the 20 selected features for training the RF algorithm to predict the S4 index are as follows: S4_mean_lag30m, S4_mean_lag25m, S4_mean_lag20m, S4_mean_lag15m, S4_mean_lag10m, S4_mean_lag5m, S4_mean_lag1m, S4_mean, cos_minute, BX_mean_3hr, F10.7_kurt_12hr, S4_mean_lag60m, F10.7_first, sin_minute, BX_mean_2hr, R_No_first, SW_Speed_first, BX_min_6hr, BY_first, and BY_first.

Table 5 demonstrates consistent improvement across all metrics following feature selection. The MAE decreased from 0.007305 to 0.007263, representing a reduction of approximately 0.58% and indicating enhanced predictive accuracy. Similarly, the R^2 score improved from 0.514377 to 0.517352, reflecting a 0.58% increase in explained variance.

Most significantly, the near-identical values of the R^2 (0.517352) and Adjusted R^2 (0.516857) scores with a minimal difference of only 0.096% provide strong evidence that the feature selection process successfully eliminated redundant variables without sacrificing explanatory power. This marks a substantial improvement over the pre-selection model, where the gap between R^2 and Adjusted R^2 was twice as large (0.194%).

The reduction of features from 40 to 20 has resulted in a more parsimonious and computationally efficient model. These findings validate the efficacy of the Gini Importance method for feature selection, confirming its ability to identify and retain only the most predictive features for the S4 storm prediction task.

4.1.2 Methodology and Analysis of S4 Time-Series Forecasting Interval Visualization

This figure 3 illustrates the model's forecasting capability by comparing the actual S4 index against the predicted mean over a continuous time interval. The visualization is constructed using the Random Forest's inherent ability to quantify uncertainty. The Predicted S4 (red dashed line) represents the average output from all trees in the ensemble. The ٪۹۵ Confidence Interval (gray shaded area) is calculated by taking plus/minus 1.96 times the standard deviation of the individual predictions across the forest's ensemble, defining a clear region of prediction uncertainty. This method allows for a robust assessment of forecast reliability, with the highly volatile Actual S4 values (blue line) demonstrating the signal's true behavior across the observed period. The x-axis plots the time of the predicted event, aligning the forecast with real-world operational scenarios.

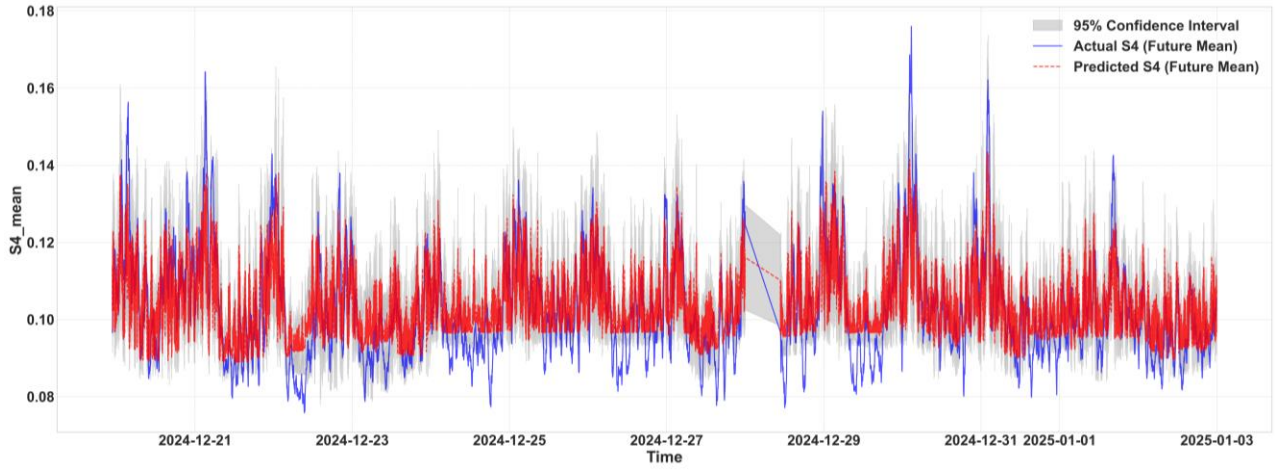

Figure 3. Time-Series Comparison of Actual vs. Predicted Future S4 Index with 95% Confidence Intervals.

As depicted in the figure, the forecast demonstrates significant strengths critical for operational space weather applications. The model's mean prediction (red line) consistently and accurately traces the complex, evolving baseline of the actual S4 index, confirming a low systematic forecasting bias. More importantly, the visualization proves the model is well-calibrated ,as the vast majority of the highly volatile actual S4 values are successfully contained within the narrow 95% Confidence Interval .This high containment rate strongly validates the reliability of the predicted uncertainty band. It is important to note the blank section observed between December 27 and 29. This is attributed to a temporary data void in the GNSS receiver network's S4 observations at that specific time, and not a limitation of the forecasting model itself.

4.2 Geomagnetic Storm Prediction

In this section, we train a Random Forest classifier to predict the presence or absence of geomagnetic storms by defining two distinct time intervals: 82, 83, 113, 114, 309, 310, 131, 132, 133, 225, 226, 282, 283, 284, 285, 1 and 2 days for storm conditions and other days for non-storm conditions. For each interval, we reserve the final 20 percent of observations as a test set, ensuring that the evaluation period ends with the event of interest, either the absence or onset of a storm. This setup guarantees that the model is both trained on representative intervals and rigorously tested on its ability to detect storm and non-storm states in real-world temporal sequences. Since this objective is binary, we require the RF Classifier method. The optimal set of hyperparameters chosen for this classification model (n_estimators=102, max_depth=3, min_samples_split=18, min_samples_leaf=18) describes a powerful yet controlled and more regularized architecture. The choice of **criterion='gini'** signifies that the model uses the Gini impurity metric, meaning that at each split, it selects the division that minimizes the probability of misclassification. Finally, **max_features='sqrt'** indicates that the number of features considered for the best split is limited to the square root of the total number of features, introducing randomness to make the model more robust and prevent overfitting.

Table 6. Features Used for Storm Presence/Absence Detection

Feature	Feature
TEC_mean	Mean of Total Electron Content (TEC).
TEC_std	Standard deviation of TEC.
TEC_max	Maximum value of TEC.
S4_mean	Mean of the S4 index.
S4_std	Standard deviation of the S4 index.
S4_max	Maximum value of the S4 index.
Kp_avg	Average of the Kp index.
Dst_min	Minimum value of the Dst index.
Dst_mean	Mean of the Dst index.
f10_7_avg	Average of the F10.7 solar flux.
BZ_mean	Mean of the BZ magnetic field component.
lat_mean	Mean latitude.
lon_mean	Mean longitude.
elevation_avg	Mean longitude.

occheight_mean	Mean occultation height.
occheight_std	Standard deviation of occultation height.
Bz_south_duration	Duration of the southward Bz component over a 2-hour
BX_mean	60-minute moving mean of the BX magnetic field component.
BY_mean	60-minute moving mean of the BY magnetic field component.
SW_Speed_mean	60-minute moving mean of the Solar Wind Speed.
R_No_mean	60-minute moving mean of the R Number (Sunspot Number proxy).
Flow_pressure_mean	60-minute moving mean of the Flow Pressure.
Proton_Density_mean	60-minute moving mean of the Proton Density.
AE_mean	60-minute moving mean of the AE index.
TEC_slope	Derivative (slope) of TEC_mean (calculated over 10 minutes).
Bz_trend	Derivative (slope) of BZ_first (calculated over 10 minutes).
TEC_second_derivative	Second derivative of TEC_mean (calculated via successive 5-minute differences).
Bz_Dst	Interaction feature: Product of BZ_mean and Dst_min (BZ_mean × Dst_min).
Kp_F10.7	Interaction feature: Product of Kp_avg and F10.7_avg (Kp_avg × F10.7_avg).
TEC_S4_ratio	Ratio of TEC_mean to S4_mean (TEC_mean / S4_mean).
Dst_drop	Difference between Dst_mean and Dst_min (proxy for Dst drop within 1 hour).
S4_ratio	Ratio of S4_std to S4_mean (S4_std / S4_mean).

Table 6 presents the features used to train the model, where storm presence is encoded as 1 and absence as 0. We evaluated performance using a confusion matrix in Fig. 4. A confusion matrix is a summary table that compares actual and predicted class labels by counting true positives, false positives, false negatives, and true negatives to evaluate classification performance. For 20% test dataset of 19177 samples, the confusion matrix shows that the model performed well overall, correctly identifying 14323 instances of 'No Storm' and 4266 instances of 'Storm', but missed 57 actual storms (False Negatives) and incorrectly flagged 531 non-storm periods as storms (False Positives).

The data reveals that the 57 missed storms constitute only approximately 1.3% of all true storm events (4323 total actual storms), resulting in an exceptionally high True Positive Rate. This low rate of False Negatives is the model's most significant strength and is far more valuable than the relatively higher number of False Positives (531 cases).

In space weather operations, the primary objective is to minimize the risk of unpredicted catastrophic events. A False Negative (a missed storm) can lead to billions of dollars in damage to satellites, communications, and power grids. Conversely, a False Positive (a false alarm) results mainly in operational inconvenience.

The model successfully prioritizes safety and preparedness by achieving a near-perfect detection rate for actual storms. By limiting the number of False Negatives to just 57, the model effectively ensures that nearly all hazardous events are successfully identified. This trade-off—accepting 531 unnecessary alerts to secure the detection of 4266 real events is precisely the desired characteristic for an effective early warning system.

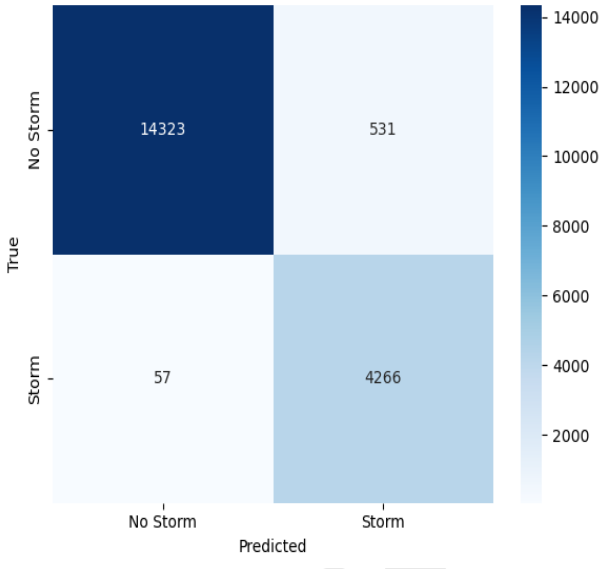

Figure 4. Confusion matrix for the test dataset.

The five metrics (Accuracy, Precision, Recall, F1-score, and Support) are presented in Table 6 to provide a more precise understanding of the machine learning model's effectiveness in predicting storms. The evaluation metrics for the test data indicate that model exhibits near-perfect performance in identifying actual storm events (Recall of 0.99 for Class 1), thereby successfully prioritizing the avoidance of False Negative errors, which is the most critical requirement for any space weather early warning system.

Table 6. Classification Report Metrics for the Model.

	Precision	Recall	F1-score	support
Non-storm (0)	1.00	0.96	0.98	14854
Storm (1)	0.89	0.99	0.94	4323

According to Figure 5, a significant point is the dependency of storm detection on the F10.7 solar index, followed by the geomagnetic indices Dst, BZ, and Kp. This shows that the model ultimately looks at the level of solar activity to predict a storm. Given that the F10.7 index reflects solar activity such as sunspots and X and M-class flares, it is not surprising that it plays a key role in predicting a storm resulting from a Coronal Mass Ejection (CME) associated with such events.

Therefore, the feature importance analysis reveals that the model correctly assigned the most weight to valid physical precursors of a storm, such as the F10.7, Dst, Kp, and BZ indices, which confirms the high potential of this approach. The model has correctly identified as most important the very features that align with our physical understanding of geomagnetic storms. Dst directly measures the ring current, which is intensely amplified during a storm's main phase; thus, Dst values on the day before a storm (during the initial phase) can provide critical predictive signals. Kp and BZ, respectively, represent the overall level of geomagnetic disturbance and the efficiency of the energy transfer mechanism from the solar wind.

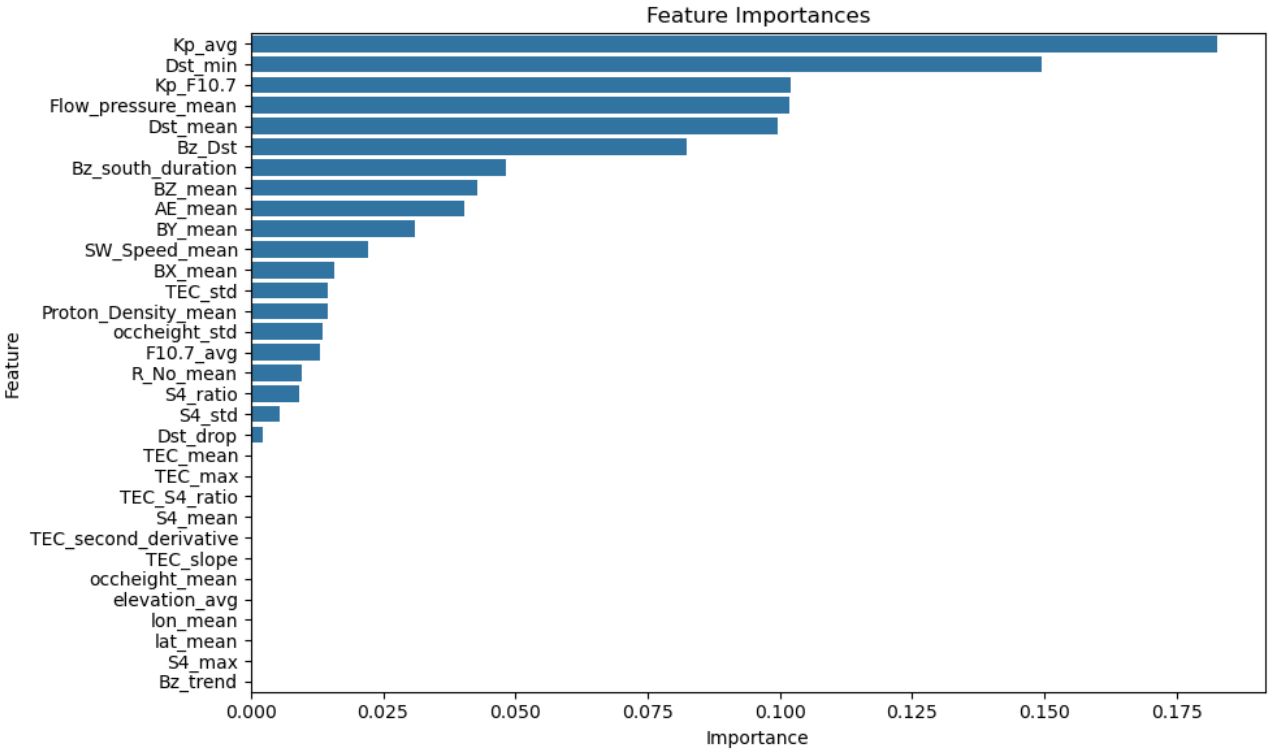

Figure 5. Feature importance ranking for the correct detection of a storm

4.2.1 Learning Curve Visualization, Performance Analysis and Discussion

This figure 6, titled "Learning Curve - Random Forest," provides a visualization of the model's performance on both training and cross-validation (CV) datasets as the volume of training examples increases. The specific metric tracked for this analysis is Accuracy .The blue line, which represents the Training Score, immediately reaches and sustains a near-perfect accuracy score of 1.0 across the entire range of sampled subsets, from the smallest starting point up to the maximum available data. In contrast, the red line, depicting the CV Score, begins at a solid baseline of approximately 0.855. It remains remarkably stable up to around 10,000 samples, before showing a modest, yet

noticeable, upward trend toward 0.88 upon incorporating the full dataset of nearly 13,000 examples. The faint red band around the CV score indicates the inherent uncertainty of the cross-validation process.

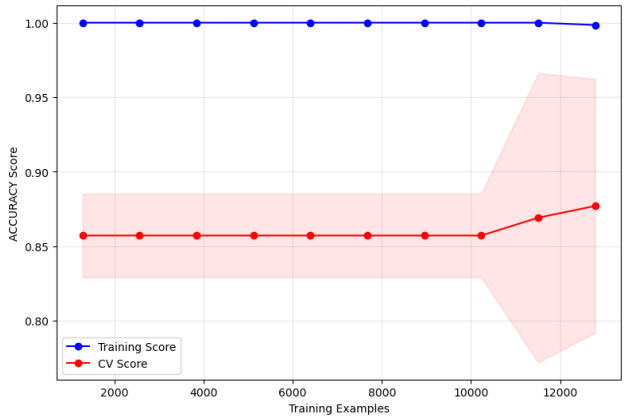

Figure 6. Learning Curve of the Random Forest Classifier for Storm Prediction.

The resulting learning curve offers compelling evidence of the Random Forest classifier's underlying robustness for this complex prediction task. Although a clear separation exists between the training and CV scores, often indicative of high variance ,the model maintains an exceptionally strong baseline CV accuracy of over 85% even when trained on minimal data. This high, consistent performance in the critical cross-validation phase immediately suggests that the model benefits from well-engineered features and possesses a fundamentally low bias .Critically, the gentle, persistent upward trajectory of the CV score as it approaches the limit of the available dataset signifies that the model is still actively learning. This positive asymptotic behavior confirms that the current system has not yet plateaued, suggesting a clear path for future improvement simply by introducing more comprehensive and diverse labeled data.

5 Conclusion

This study successfully developed and evaluated two distinct RF models to address critical challenges in space weather forecasting: the prediction of the S4 amplitude scintillation index and the binary classification of geomagnetic storm occurrence. The research demonstrates the capability of ML, specifically the RF algorithm, to capture the complex, non-linear relationships between interplanetary and geomagnetic parameters and their ionospheric impacts.

The first objective was to predict the S4 amplitude scintillation index using a robust set of 40 features derived from solar wind, geomagnetic indices, and historical S4 statistics. The initial RF regression model demonstrated moderate predictive capability. The R^2 score of 0.514 and its closely aligned Adjusted R^2 score of 0.513 indicated that while the model captured a portion of the data's variance, there was substantial room for improvement in its explanatory power.

To enhance model efficiency and generalizability, Gini Importance was employed, which systematically pruned redundant features. The refined model, trained on an optimal set of 20 features, yielded superior results: consistent improvements across all error metrics (MAE, MSE, RMSE) and a key improvement in the R^2 score to 0.517. Most notably, the gap between R^2 and Adjusted R^2 became virtually negligible (0.096%), confirming a more robust and balanced model. This demonstrates that Gini Importance effectively streamlined the model, enhancing its statistical efficiency without sacrificing predictive accuracy. Feature importance analysis correctly identified key drivers, with the lags of S4 (S4_mean_lags) and geomagnetic indices like Bx and F10.7 being among the most critical predictors.

The second objective involved training an RF classifier to distinguish between storm and non-storm conditions. The five metrics (Accuracy, Precision, Recall, F1-score, and Support) are presented in Table 6 to provide a more precise understanding of the machine learning model's effectiveness in predicting storms. The evaluation metrics for the test data indicate that the model possesses a highly reliable operational characteristic, prioritizing the detection of all actual storm events over the strict avoidance of false alarms. The storm prediction model exhibits a strong and strategically focused operational performance prioritizing high sensitivity to hazardous events. The model achieved a flawless Precision score of 1.00 for the 'Non-storm' class, meaning it never mistakenly classified a true storm as a non-storm event (zero False Omissions). The primary strength, however, lies in its near-perfect Recall of 0.99 for the 'Storm' class, indicating it successfully detected

99% of all actual storms and only missed approximately 1% of them. In critical space weather applications, this robust avoidance of False Negatives is paramount, as the cost of a single missed event far outweighs that of a false alert. The model maintains a strong Precision of 0.89 for storm predictions, meaning 89% of its alerts are correct, demonstrating that the acceptance of an 11% False Positive rate is a calculated trade-off to ensure maximum safety and reliability. Overall, the model is highly sensitive to storm conditions, and its balanced F1-scores confirm its reliability across both common and rare classification outcomes. Crucially, the feature importance analysis validated the model's physical interpretability, as it correctly assigned the highest weights to well-established storm precursors: the F10.7 solar flux, followed by the Dst, BZ, and Kp indices.

6 REFERENCES

[1] Taran, Somayeh, Nasibe Alipour, Kourosh Rokni, S. Hadi Hosseini, Omid Shekoofa, and Hossein Safari. "Effect of geomagnetic storms on a power network at mid latitudes." Advances in Space Research 71, no. 12 (2023): 5453-5465.

[2] Saiz, Elena, Yolanda Cerrato, Consuelo Cid, Venera Dobrica, Pavel Hejda, Petko Nenovski, Peter Stauning et al. "Geomagnetic response to solar and interplanetary disturbances." Journal of Space Weather and Space Climate 3 (2013): A26.

[3] S. Priyadarshi, "A review of ionospheric scintillation models," Surveys in geophysics, vol. 36, no. 2, pp. 295–324, 2015.

[4] J. Vil` a-Valls, P. Closas, C. Fern´andez-Prades, and J. T. Curran, "On the mitigation of ionospheric scintillation in advanced gnss receivers," IEEE Trans. Aerosp. Electron. Syst., vol. 54, no. 4, pp. 1692–1708, 2018.

[5] DeFranceschi, G.; Spogli, L.; Alfonsi, L.; Romano, V.; Cesaroni, C.; Hunstad, I. The ionospheric irregularities climatology over Svalbard from solar cycle 23. Sci. Rep. 2019, 9, 9232.

[6] Sreeja, V. "Impact and mitigation of space weather effects on GNSS receiver performance." Geoscience letters 3, no. 1 (2016): 24.

[7] ESA, "Space weather study results," 2016. [Online]. Available: https://esamultimedia.esa.int/docs/business with esa/ Space Weather Cost Benefit Analysis ESA 2016.pdf

[8] M. A. Kelly, J. M. Comberiate, E. S. Miller, and L. J. Paxton, "Progress toward forecasting of space weather effects on uhf satcom after operation anaconda," Space Weather, vol. 12, no. 10, pp. 601–611, 2014.

[9] T. E. Humphreys, M. L. Psiaki, B. M. Ledvina, A. P. Cerruti, and P. M. Kintner, "Data-driven testbed for evaluating gps carrier tracking loops in ionospheric scintillation," IEEE Trans. Aerosp. Electron. Syst., vol. 46, no. 4, pp. 1609–1623, 2010.

[10] Y. Jiao, D. Xu, C. L. Rino, Y. T. Morton, and C. S. Carrano, "A multifrequency gps signal strong equatorial ionospheric scintil lation simulator: algorithm, performance, and characterization," IEEE Trans. Aerosp. Electron. Syst., vol. 54, no. 4, pp. 1947 1965, 2018.

[11] De Oliveira Moraes, A.; Costa, E.; De Paula, E.R.; Perrella, W.J.; Monico, J.F.G. Extended ionospheric amplitude scintillation model for GPS receivers. Radio Sci. 2014, 49, 315–329.

[12] Klobuchar, J.A. Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron. Syst. 1987, 3, 325–331.

[13] Prieto-Cerdeira, R.; Orús-Pérez, R.; Breeuwer, E.; Lucas-Rodriguez, R.; Falcone, M. Performance of the Galileo single-frequency ionospheric correction during in-orbit validation. GPS World 2014, 25, 53–58.

[14] Yuan, Y.; Wang, N.; Li, Z.; Huo, X. The BeiDou global broadcast ionospheric delay correction model (BDGIM) and its preliminary performance evaluation results. Navigation 2019, 66, 55–69.

[15] S. Basu, K. Groves, S. Basu, and P. Sultan, "Specification and forecasting of scintillations in communication/navigation links: current status and future plans," Journal of atmospheric and solar-terrestrial physics, vol. 64, no. 16, pp. 1745–1754, 2002.

[16] Groves, K.M.; Basu, S.; Weber, E.J.; Smitham, M.; Kuenzler, H.; Valladares, C.E.; Sheehan, R.; MacKenzie, E.; Secan, J.A.; Ning, P.; et al. Equatorial scintillation and systems support. Radio Sci. 1997, 32, 2047–2064.

[17] Aol, S.; Buchert, S.; Jurua, E. Ionospheric irregularities and scintillations: A direct comparison of in situ density observations with ground-based L-band receivers. Earth Planets Space 2020, 72, 164.

[18] Pappoe,J.A.; Yoshikawa, A.; Kandil, A.; Mahrous, A. A machine learning approach combined with wavelet analysis for automatic detection of Pc5 geomagnetic pulsations observed at geostationary orbits. Adv. Space Res. 2023, in press.

[19] Pappoe, J.A.; Yoshikawa, A.; Kandil, A.; Mahrous, A. Machine learning techniques for estimation of Pc5 geomagnetic pulsations observed at geostationary orbits during solar cycle 23. J. Atmos. Sol. Terr. Phys. 2024, 260, 106258.

[20] Farooki, H.; Abduallah, Y.; Noh, S.J.; Kim, H.; Bizos, G.; Shin, Y.; Wang, J.T.L.; Wang, H. A Machine Learning Approach to Understanding the Physical Properties of Magnetic Flux Ropes in the Solar Wind at 1 au. Astrophys. J. 2024, 961, 81.

[21] Halawa, S. S., Alansaari, M. A., Sharif, M. E., Alhammadi, A. M., & Fernini, I. (2024, July). Ionospheric Scintillation Forecasting Using Machine Learning. In IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium (pp. 5635-5639). IEEE.

[22] Nasurudiin, S., Yoshikawa, A., Elsaid, A., & Mahrous, A. (2024). Prediction of Ionospheric Scintillations Using Machine Learning Techniques during Solar Cycle 24 across the Equatorial Anomaly. Atmosphere, 15(10), 1213.

[23] Zhan, Y. H., Xing, Z. Y., Xu, T., Zhang, Q. H., Oiao, F., & Wang, Y. (2024, October). Prediction of Northern Polar ionospheric scintillation Using the RF and LSTM Machine Learning Methods. In 2024 14th International Symposium on Antennas, Propagation and EM Theory (ISAPE) (pp.

1-4). IEEE.

[24] Trachuentong, S., Supnithi, P., Myint, L. M. M., & Saito, S. (2024, May). Ionospheric Scintillation Prediction Using Decision Tree and Rainforest Techniques. In 2024 10th International Conference on Engineering, Applied Sciences, and Technology (ICEAST) (pp. 89-92). IEEE.

[25] Dittmann, T., Morton, Y. J., & Chang, H. (2025, January). Machine Learning Classification of Ionosphere and RFI Disturbances in Spaceborne GNSS GNSS-RO Measurements. In Proceedings of the 2025 International Technical Meeting of The Institute of Navigation (pp. 161-176).

[26] Kuruva, L., Avula, M. R., & Achanta, D. S. (2024, July). ML based model for Detection of Ionospheric Scintillations using PolaRx5S data. In 2024 IEEE Space, Aerospace and Defence Conference (SPACE) (pp. 678-681). IEEE.

[27] Das, A.; Das Gupta, A.; Ray, S. Characteristics of L-band (1.5 GHz) and VHF (244 MHz) amplitude scintillations recorded at Kolkata during 1996–2006 and development of models for the occurrence probability of scintillations using neural network. *J. Atmos. Sol. Terr. Phys.* 2010, *72*, 685–704.

[28] Rezende, L.F.C.; de Paula, E.R.; Stephany, S.; Kantor, I.J.; Muella, M.T.A.H.; de Siqueira, P.M.; Correa, K.S. Survey and prediction of the ion-ospheric scintillation using data mining techniques. *Space Weather* 2010, *8*, 1–10.

[29] Nasurudiin, S.; Yoshikawa, A.; Elsaid, A.; Mahrous, A. Prediction of Ionospheric Scintillations Using Machine Learning Techniques during Solar Cycle 24 across the Equatorial Anomaly. *Atmosphere* 2024, *15*, 1213.

[30] O. Carvalho, P. A. A. da Silva de Almeida Nava Alves, R. Y. de La Cruz Cueva and A. O. B. Filho, "Nowcasting of Amplitude Ionospheric Scintillation Based on Machine Learning Techniques," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 6, pp. 4917-4927, Dec. 2022, doi: 10.1109/TAES.2022.3188741.

[31] A. M. Darya, A. A. Al-Owais, M. M. Shaikh and I. Fernini, "Amplitude Scintillation Forecasting Using Bagged Trees," *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 2022, pp. 2275-2278, doi: 10.1109/IGARSS46834.2022.9883380.

[32] Atabati, A.; Jazireeyan, I.; Alizadeh, M.; Langley, R.B. Prediction of Ionospheric Scintillation with ConvGRU Networks Using GNSS Ground-Based Data across South America. *Remote Sens.* 2024, *16*, 2757.

[33]Tian, P., Yu, B., Ye, H., Xue, X., Wu, J., & Chen, T. (2022). Estimation model of global ionospheric irregularities: An artificial intelligence approach. Space Weather, 20, e2022SW003160.

[34]Tian, P., Yu, B., Ye, H., Xue, X., Wu, J., & Chen, T. (2024). Deep learning insights into ionospheric Sporadic E irregularities under different solar activity conditions. Journal of Geophysical Research: Machine Learning and Computation, 1, e2024JH000279.

[35] Biau, Gérard, and Erwan Scornet. "A RF guided tour." Test 25, no. 2 (2016): 197-227.

[36] Büyükkeçeci, M., & Okur, M. C. (2023). A comprehensive review of feature selection and feature selection stability in machine learning. Gazi University Journal of Science, 36(4), 1506-1520.

[37] Mueller, T., Segin, A., Weigand, C., & Schmitt, R. H. (2023). Feature selection for measurement models. International journal of quality & reliability management, 40(3), 777-800.

[38] Dunne R, Reguant R, Ramarao-Milne P, Szul P, Sng LMF, Lundberg M, Twine NA, Bauer DC. Thresholding Gini variable importance with a single-trained random forest: An empirical Bayes approach. Comput Struct Biotechnol J. 2023 Sep 1;21:4354-4360. doi: 10.1016/j.csbj.2023.08.033. PMID: 37711185; PMCID: PMC10497997.

[39] Pardeshi, Nilesh & Patil, Dipak. (2023). Applying Gini Importance and RFE Methods for Feature Selection in Shallow Learning Models for Implementing Effective Intrusion Detection System. 10.2991/978-94-6463-136-4_21.