

# Machine Learning Algorithm for the Influence of Space Weather Events of Geomagnetic Activity on GNSS Services

Lucija Žužić<sup>1, 2</sup>, Deni Klen<sup>1, 2</sup>, Darko Špoljar<sup>3</sup> and Renato Filjar<sup>1, 2, 4 \*</sup>

<sup>1</sup> Department of Computer Engineering, Faculty of Engineering, University of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia

<sup>2</sup> Center for Artificial Intelligence and Cybersecurity, University of Rijeka, Radmile Matejcic 2, 51000 Rijeka, Croatia

<sup>3</sup> Department for Transport Logistics, Hrvatsko Zagorje Krapina University of Applied Sciences, Setaliste hrvatskog narodnog preporoda 6, 49000 Krapina, Croatia

<sup>4</sup> Department for Informatics, Hrvatsko Zagorje Krapina University of Applied Sciences, Setaliste hrvatskog narodnog preporoda 6, 49000 Krapina, Croatia

\*E-mail of corresponding author: renato.filjar@uniri.hr

**Abstract:** This study proposes a machine learning-based model that utilises the Disturbance storm-time (*Dst*) index to categorise space weather events in sub-equatorial regions and assess their impact on Global Navigation Satellite System (GNSS) Positioning, Navigation, and Timing (PNT) performance. Performance was benchmarked using various indicators derived from a confusion matrix for each *Dst* class defined in the study. Several algorithms from various families, including probabilistic methods, tree-based methods, ensemble learning methods, and a deep neural network (NN), using Long Short-Term Memory (LSTM) and bidirectional or convolutional layers, were evaluated on historical *Dst*-index-labelled GNSS data. The LSTM and convolutional-layer methods achieved the highest accuracy, but other methods had much lower inference times. The study presents a novel framework for classifying the impact of geomagnetic activity on GNSS services, thereby contributing to the development of an early-warning system for PNT degradation and to the establishment of a tailored GNSS ionospheric correction model for space weather conditions.

**Keywords:** Global Navigation Satellite System Positioning, Navigation, and Timing, space weather conditions, machine learning, classification model, Total Electron Content, Disturbance storm-time.

**Received** 10 February 2026

**Accepted** 18 March 2026

**Published** 6 April 2026

## 1. Introduction

The Global Navigation Satellite System (GNSS) and its Positioning, Navigation, and Timing (PNT) service have matured into an essential part of national infrastructure, public goods, and are the enablers of a vast number of emerging technologies and socio-economic applications [1, 2]. Maintaining GNSS PNT quality is crucial for the sustainable development of the modern economy and society [2, 3]. Overcoming the shortcomings and vulnerabilities of GNSS PNT poses a scientific challenge [4, 5, 6] and is necessary for a diverse range of scientists, developers, operators, regulators, and users of GNSS-based systems and services [2]. Modifications and novel approaches to PNT algorithms, information fusion, and supporting computing and communications architectures have been under development to make GNSS

PNT resilient against the natural and artificial adversarial effects [2, 6]. Those include introducing GNSS positioning environment (ambient) awareness and alignment to GNSS applications to meet their needs and demands for the PNT performance [6].

The integrity of GNSS services is increasingly threatened by adverse space weather conditions, particularly geomagnetic storms [4, 7, 8]. The morphology of these intense geomagnetic storms [6, 9] can be studied using geophysical data analysis based on geomagnetic observatory data, solar wind flux speed, and interplanetary magnetic field parameters [7, 8]. The Disturbance storm-time (*Dst*) index is a widely accepted geomagnetic activity indicator, used to evaluate the intensity of such storms [4, 7, 8]. The geomagnetic activity, including pulsations, is strongly modulated by storm conditions, local time, and magnetic latitude. Understanding the magnetosphere-ionosphere coupling, introduced and described in [10], is important for space weather prediction. Space weather, geomagnetic, and ionospheric conditions may modify the GNSS signal characteristics, thus leading to significant PNT degradation, especially in safety-critical applications



© The Author(s). Published by CIEES Association Press. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Generic License.

such as aviation and maritime navigation. Recent developments in artificial intelligence have enabled the classification and prediction of space weather conditions using machine learning (ML) methods [10, 11, 12].

The Earth's ionosphere, a part of the Earth's atmosphere stretching from 50 km to 2000 km above the Earth's surface and comprised of rare but mostly charged atoms and molecules, is the major natural cause of GNSS PNT degradation [7]. This phenomenon especially affects most currently used GNSS receivers, which operate as single-frequency receivers and are exposed to GNSS ionospheric effects [1]. Driven by powerful and random flows of energy and particles from the Sun (space weather conditions), the ionospheric conditions define the properties of GNSS signal propagation through the Earth's atmosphere and the resulting ionospheric delay [7]. The GNSS ionospheric delay introduces errors in the GNSS PNT algorithm, which is designed to produce position, velocity, and time estimates for a GNSS receiver [1, 3]. The ionosphere affects GNSS satellite signals for position estimation by introducing signal propagation delay [1, 3]. The GNSS ionospheric delay is a stochastic variable whose value is determined by complex physical processes of space weather [1, 3]. The GNSS signal encounters a certain number of charged particles from the satellite to a mobile unit (GNSS receiver) [3]. This encounter is quantified using the Total Electron Content (*TEC*) defined by (1):

$$TEC = \int_{\text{upper ionospheric boundary}}^{\text{lower ionospheric boundary}} N(h)dh \quad (1)$$

in electrons/m<sup>2</sup> [1, 3, 7], where  $h$  denotes the height above the Earth's mean sea level in m,  $N(h)$  represents the vertical ionospheric profile, the volume density of charged particles at height  $h$ , in electrons/m<sup>2</sup>. For convenience, *TEC* may be expressed in *TEC* Units (TECU), with 1 TECU = 1016 electrons/m<sup>2</sup>. The *TEC* dataset used in this study was derived from Receiver Independent Exchange Format (RINEX) *TEC* observations using Global Positioning System (GPS) *TEC* software by Seemala [13].

It should be noted that in the sense of (1), *TEC* is defined as a result, a consequence, of the ionospheric conditions, and not their descriptor [5, 6, 11]. The GNSS ionospheric delay [10] may be determined by derivation from the Appleton-Hartree equation, as given in (2):

$$\Delta t_{iono} = \frac{40.3}{cf^2} \int_{\text{upper ionospheric boundary}}^{\text{lower ionospheric boundary}} N(h)dh \quad (2)$$

where  $\Delta t_{iono}$  denotes the GNSS ionospheric delay in s,  $c$  denotes the velocity of an electromagnetic wave in vacuum in m/s, and  $f$  denotes the carrier wave frequency of the satellite signal in Hz [1, 3].

Combining (1) and (2), one can conclude the linear relation between *TEC* and  $\Delta t_{iono}$ , as given in [1] and (3):

$$\Delta t_{iono} = \frac{40.3}{cf^2} TEC \quad (3)$$

The GNSS ionospheric delay has been identified as a source of GNSS PNT degradation since the dawn of GNSS. GNSS systems offer various standard GNSS ionospheric delay estimation (correction) models to mitigate the deteriorating effects on GNSS PNT, such as the Klobuchar model [1, 8]. The standard ionospheric correction models are global and insufficiently flexible to update, thereby mitigating GNSS ionospheric delay and satisfying the rising GNSS PNT performance demands [1].

The development of regional and local models, such as the one presented in this paper, aims to address the issue of GNSS PNT sustainable performance under various ionospheric conditions. Understanding the long-term non-linear association between geomagnetic disturbances and GNSS positioning errors is a key step in developing advanced correction procedures to mitigate these environmental effects. The Ambient-Aware Application-Aligned (abbreviated as (AA)<sup>2</sup>) PNT has been introduced recently to account for the actual ionospheric and geomagnetic conditions in the immediate vicinity of a mobile unit (a GNSS receiver) [6]. Direct measurements of the immediate geomagnetic and ionospheric condition variables can be supplied to a machine-learning-based, adapted GNSS ionospheric correction model, thereby providing the framework for solving the single-frequency GNSS problem [6]. Previous research has identified predictors and target variables (descriptors of geomagnetic, ionospheric, and GNSS PNT conditions). Current space weather severity scales, such as those provided by the National Oceanic and Atmospheric Administration (NOAA), are based on global space weather and geomagnetic indices averaged over a given period, such as 3 hours for the global  $K_p$  index [11].

This study describes a methodology for machine learning-based classification of ionospheric conditions, which is entirely based on observations of geomagnetic indices. The model is sufficiently simple to be applied on computationally capable platforms with suitable geomagnetic field sensors, such as smartphones and connected/autonomous vehicles. The research presented acquires ambient data and examines its statistical properties. The dataset is split into training and test sets. Several candidate GNSS ionospheric delay models are developed for *Dst* data from the INTERMAGNET [14] dataset reformatted to match the *TEC* data format. The ML methods represent examples from different families, including probabilistic methods (a linear model (LM)), tree-based methods implemented in the "rpart" package, ensemble learning methods (generalised additive model boosting (GAMB), bagged classification and regression trees (BCART), and gradient boosting machine (GBM)), and deep neural network (NN) models using Long Short-Term Memory (LSTM) and bidirectional or convolutional layers. The proposed deep learning LSTM architecture is an original contribution by the authors, redesigned for multiple classes in a different research area, and represents a significant extension of the original architecture's previous application in binary peptide classification.

ML approaches in space weather classification generally fall into two categories: traditional statistical models and deep learning architectures. Traditional methods, such as LMs and tree-based models, offer high interpretability and computational efficiency but may struggle to capture complex non-linearities in geophysical data. Conversely, deep learning models, particularly Recurrent Neural Network (RNN) models and LSTM networks, excel at modelling temporal dependencies in sequential data, making them well-suited for characterising the evolution of geomagnetic storms.

A tailored set of validation methods is used to assess their performance. The optimal GNSS ionospheric delay

correction model is identified based on GNSS PNT-related objective criteria, and its performance is demonstrated in an independent case study. It is important to clarify that this framework is designed for nowcasting and classification rather than forecasting future indices. By classifying the current storm phase based on real-time indices, the system serves as a trigger for adaptive GNSS algorithms. Upon detecting an extreme or recovery phase, the PNT engine can immediately switch to a specialised ionospheric correction model or re-weight satellite observations, thereby mitigating PNT degradation during the event.

## 2. Method

A linear model (LM), tree-based methods from the "rpart" package, ensemble learning methods (generalised additive model boosting (GAMB), bagged classification and regression trees (BCART), and gradient boosting machine (GBM)), and custom deep neural network (NN) models using Long Short-Term Memory (LSTM) and bidirectional or convolutional layers [15, 16, 17] were tested based on their ability to classify a set of observations [18] of the geomagnetic field in *TEC*, and other predictors, into one of the scenarios of space weather conditions based on *Dst*. The methods were selected because they represent larger families [15, 16, 17] of machine learning methods [18]. Where appropriate, simpler linear models may be preferable unless deep learning models substantially improve predictive performance. This is why simpler models were benchmarked alongside computationally intensive LSTM models. The full data and code are on Zenodo (<https://doi.org/10.5281/zenodo.14502657>, accessed on 3 September 2025).

All tested models, except the LSTM deep learning models, were applied as described in the research by Kuhn for the "caret" package [19], which was used in R [20]. The deep learning models were implemented in Python version 3.10.13 using the "keras" library version 2.10.0 to replicate the authors' previous research, which used a similar model architecture in the domain of peptide self-assembly [21].

Cross-validation can be used to detect and potentially reduce overfitting problems with statistical methods [21]. The "trainControl" function was used to tune the hyperparameters of the non-deep learning models implemented in the R environment [20] using the "caret" package [19] in k-fold cross-validation with 10 folds and repeated 10 times.

To measure inference time and training time, the experiment was run on the 6<sup>th</sup> Generation Intel(R) Core(TM) i5-6300HQ Central Processing Unit (CPU) of a personal computer, running at a base speed of 2.3 GHz, instead of a Graphics Processing Unit (GPU), to better estimate the execution time for mobile devices and built-in systems using GNSS PNT that have significantly lower computational capabilities.

An algorithm (block diagram) of the method is presented in Figure 1.

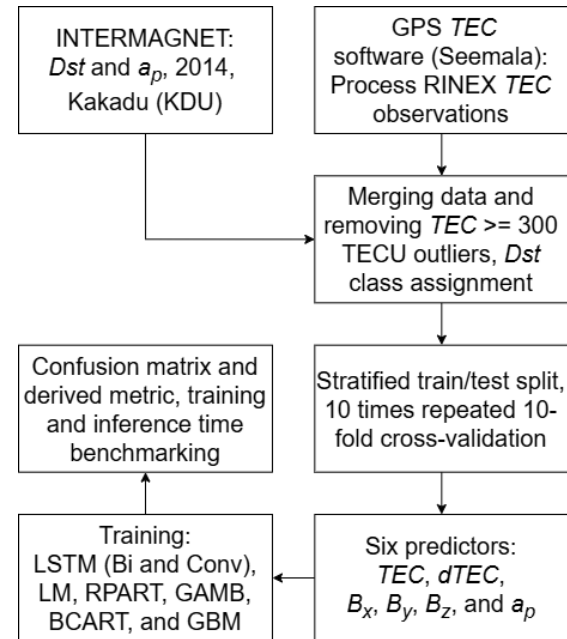


Fig. 1. The full method workflow from data processing to ML model training and testing.

GPS *TEC* software by Seemala was used to obtain RINEX *TEC* observations [13], and this data was merged with the INTERMAGNET [14] dataset, which included *Dst* and  $a_p$  data from 2014 for a measuring station maintained by Geoscience Australia in Kakadu (KDU). The  $TEC \geq 300$  TECU outliers were removed before *Dst* classes were determined. The training and testing datasets were stratified by class, and the models were evaluated using 10-fold cross-validation (repeated 10 times) [18, 19, 20]. Six predictors were used in the LSTM (bidirectional (Bi) and convolutional (Conv)), LM, RPART, GAMB, BCART, and GBM models for training [15, 16, 17]. Results evaluation included confusion matrix performance indicators and comparisons of inference and training times.

### 2.1. Linear models

The linear model (LM) assumes linearity [15, 19, 20]. A substantial reduction in the complexity of the related statistical theory is possible by using least squares analysis, such as the minimum of a sum of squares function [16, 17, 18].

### 2.2. Tree models

A decision tree (DT) is used for supervised learning in statistics and machine learning [15, 16, 17]. A tree is recursively partitioned by dividing the original set (the root node) into subsets that form descendants (successors), using classification rules based on features [18]. Random forests (RFs) are ensemble models that implement bagging over a set of DT classifiers. This effectively decreases overfitting. Multiple trees with different feature subsets [22] and training samples are created to avoid depending on a single tree. The final decision is made on the majority classification of these trees.

### 2.3. Generalised additive model boosting

In a generalised additive model (GAM), the response variable depends linearly on the smooth functions of one or more predictor variables, combining the properties of generalised linear models and additive models. GAMs can be interpreted as the discriminative generalisation of the naive Bayes generative model [15, 16, 17]. Parametric, nonparametric, or semiparametric function forms can be used to improve fit [18]. These methods include the risk of losing interpretability [15, 16, 17]. GAM Boosting (GAMB) fits a generalised additive model using likelihood-based boosting [18].

### 2.4. Bagged tree models

Bootstrap aggregating, or bagging, is an ensemble meta-algorithm that increases the stability and accuracy of ML classification and regression algorithms by combining the predictions of randomly generated training sets [15, 16, 17]. Many weak learners typically aggregate to outperform a single learner across the entire set and tend to overfit less [18]. Training for individual weak learners can be performed in parallel, as each separate bootstrap can be processed independently.

### 2.5. Gradient boosting machines

A gradient boosting machine (GBM) is an algorithm based on boosting in a functional space, where the target is pseudo-residuals rather than residuals, as in traditional boosting [15, 16, 17]. The idea of gradient boosting originated from the observation that boosting can be interpreted as an optimisation algorithm [18].

### 2.6. LSTM bidirectional and convolutional models

The LSTM architecture conceived by the authors for binary peptide classification [21] has been successfully adapted and extended for *Dst*-index multiclass classification. Inspired by peptide [21] feature scaling, the variable values were again scaled to the [0, 1] range, facilitating gradient flow and yielding a model with superior performance [23]. Since the number of predictors was uniform, a masking layer and padding for processing varying peptide sequence lengths were unnecessary. Unlike in previous work on peptide classification [21], there were no sequences of multiple values for the same feature, such as AP (Aggregation Propensity) of amino acids, dipeptides, and tripeptides, so a concatenation layer was omitted. One version of the tested models applied two LSTM layers with 5 units to the input data, with the first layer being bidirectional. These submodels also encompassed a dense layer with 64 units, as obtained by hyperparameter optimisation in previous work [21], followed by the scaled exponential linear unit (SELU) activation. Other tested models utilised two one-dimensional convolutional layers, with 5 filters and a kernel size of 4, as determined through hyperparameter tuning in the preceding experiments on peptides [21]. The number of filters was inspired by previous research on therapeutic peptides [24]. The training was limited to 5 epochs [25], and the batch size was set to 600 for the fastest operation and smoother gradients. The initial learning rate was set to 0.01. Only the model with the lowest training

loss was saved. A 50% dropout regularisation was applied to the final layer of all submodels to prevent overfitting. The sigmoid activation function was applied to the final layer of each model to yield a value between 0 and 1, representing the probability that an input sequence exhibits self-assembly in the original model. In the *Dst* model, this value is converted to the original variable range, with 0 representing the minimum and 1 representing the maximum. Values are then rounded to the nearest integer to obtain the class number.

### 2.7. Confusion matrix and statistical test specifications

All tested candidate models were compared to the ground truth. A confusion matrix and derived metrics were used, including the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) classifications, accuracy, true positive rate (TPR) or sensitivity, true negative rate (TNR) or specificity, *F1* score, balanced accuracy (BA), positive predictive value (PPV), and negative predictive value (NPV) [15, 16, 17]. These metrics [18] and the 95% confidence interval (CI) were calculated using the "epi.tests" function from the "epiR" package developed for R [20]. The parameters "conf.level" and "digits" were 0.95 and 4, respectively. The inputs were the TP, TN, FP, and FN values shown in confusion matrix heatmaps in Figure 2 in further text for each tested model.

## 3. Data

Dynamic space weather conditions, such as solar activity and geomagnetic storms, can impact GNSS PNT performance and affect high-frequency GPS signals passing through the ionosphere, motivating research on error modelling. Geomagnetic storms degrade signal quality by altering the variability of the Global Electric Current (GEC). The ionosphere may show changes related to location, geomagnetic and solar activity, sunspots, local time, seasonality, thunderstorms, nuclear experiments, earthquakes, and other phenomena. This study focuses on parameters describing disturbances of the Earth's geomagnetic field, most importantly  $a_p$ -indices derived from  $K_p$ -indices, which are calculated using  $K$ -indices, the Total Electron Content (*TEC*), the standard deviation of the Total Electron Content (*dTEC*), and Disturbance storm-time (*Dst*). Incorporating parameters such as the  $K_p$ -indices and  $a_p$ -indices, which provide global measures of geomagnetic activity, alongside local *TEC* and *Dst*-index values, enables a more detailed assessment of the space environment and its potential effects on GNSS signals. Values of  $a_p$ , *TEC*, *dTEC*, and *Dst* were used with the  $B_x$ ,  $B_y$ , and  $B_z$  components of the Earth's magnetic field to train machine-learning models. The *dTEC* was analysed based on all measurements in each *Dst* class and derived from the RINEX observations within the hourly window.

Multiple *Dst*-dependent classes were predefined using theoretical knowledge. Statistical analysis of the data confirmed that the distributions of other variables differ across different *Dst* ranges, supporting the validity of the classification. The study assumes that the dependent output variable, the *Dst* class, can be predicted from the independent input variables. *TEC* data were obtained using GPS *TEC* software by Seemala to process RINEX *TEC*

observations [13]. The INTERMAGNET [14] dataset contains  $Dst$  and  $a_p$  data from 2014 for a measuring station maintained by Geoscience Australia in Kakadu, referred to as KDU in the database, at 12.69 degrees south latitude and 132.47 degrees east longitude, near Darwin, Northern Territory, Australia. The two datasets are merged based on location, year, month, day, and time of day (hours and minutes).

### 3.1. Magnetic field indices

The Earth's magnetic field resembles that of a bar magnet. However, plasma emanating from the solar corona and the Sun's domain both influence the interplanetary magnetic field. The  $B_x$ ,  $B_y$ , and  $B_z$  vectors represent the components of the interplanetary magnetic field.  $B_x$  and  $B_y$  are parallel to the plane of orbits, and the third component  $B_z$  is perpendicular.

### 3.2. Geomagnetic storm indices

The geomagnetic storm  $K$ -index is an integer from 0 to 9 measuring disturbances in global geomagnetic activity. The maximum positive and negative fluctuations of the horizontal components of the Earth's magnetic field,  $B_x$  and  $B_y$ , over 3 hours relative to a quiet day are added to determine the total maximum fluctuation. Each observatory uses different threshold values to convert the maximum nT (nano-Tesla) fluctuation to a  $K$ -index value. The thresholds for each observatory are adjusted so that the historical occurrence rates for each  $K$ -index value are similar across all observatories. Observatories with a lower geomagnetic latitude use a lower fluctuation in  $B_x$  and  $B_y$  to achieve each  $K$ -index value.

### 3.3. Planetary geomagnetic storm indices

The planetary geomagnetic storm  $K_p$ -index is derived from 3-hour-based  $K$ -indices from 13 magnetometer stations between 44 and 60 degrees of north and south latitude. Hourly geomagnetic storm index data are available on the National Aeronautics and Space Administration (NASA) Goddard Space Physics Data Facility (SPDF) web pages. The scale values of the  $K_p$ -index are determined by the change of the geomagnetic field and the geomagnetic storm effect in nT. The official planetary  $K_p$ -index is a weighted average of multiple observatories. When  $K$ -index data is not available in real-time, operators such as the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) calculate near-real-time estimates of the  $K_p$ -index [26].

### 3.4. Equivalent three-hourly range geomagnetic storm indices

The  $A$ -index represents a daily average level of magnetic activity. Because the relationship between the  $K$ -index and magnetometer fluctuations is nonlinear, the  $K$ -index values are not directly used for calculating average values. Each  $K$ -index or  $K_p$ -index is converted to the "equivalent three-hourly range"  $a$ -index or  $a_p$ -index, which uses a linear scale. An average of 8  $a$ -indices (lowercase) is used as the daily  $A$ -index (uppercase).

### 3.5. Disturbance storm-time

Disturbance storm-time ( $Dst$ ), also known as the geomagnetic activity  $Dst$  Index, represents an averaged measure of geomagnetic storm intensity in the Earth's subequatorial region. The  $Dst$ -index is obtained by post-processing, with the final version often published months after experimental observations were collected. It clearly shows the progression of geomagnetic disturbances at various levels and, consequently, provides invaluable input for assessing GNSS PNT performance degradation due to space weather events. The  $Dst$ -index is a geomagnetic indicator of magnetic flux changes derived from measurements taken by a network of ground-based magnetometer stations near the magnetic equator, which continuously monitor  $B_x$  and  $B_y$ , the horizontal components of Earth's magnetic field [27]. The  $Dst$ -index describes ring currents that form above the sub-equatorial region and affect the ionospheric regions in mid-latitudes. To calculate the  $Dst$ -index, variations in  $B_x$  and  $B_y$ , the horizontal magnetic field, are obtained from multiple stations and averaged. The average is subtracted from a baseline value representing the quiet-time magnetic field. The resulting  $Dst$  value in nT measures the intensity of geomagnetic disturbances, with increasingly negative values indicating stronger geomagnetic storms. The  $Dst$ -index measurements (as an hourly average) were evaluated and published on a web interface through a collaboration between NASA Goddard SPDF and the Geomagnetism and Space Magnetism Data Analysis Center of the Institute of Science, Kyoto University, Japan. Loewe and Pröls [28] classified magnetic activity  $Dst$ -indices into 5 storm classes in 1997. Gonzalez et al. [29] used 3 groups for the same data in 1994, similar to Kamide et al. in 1998 [30], Rozhnoi et al. in 2004 [31], and Contadakis et al. in 2012 [32].

### 3.6. Training, testing, and validation dataset

The original set of observations has been split into training and test subsets while sustaining proportions of data classes for geomagnetic disturbances (quiet/normal geomagnetic conditions, the positive phase of a geomagnetic storm, deep negative depression of a geomagnetic storm, and the negative recovery phase of a geomagnetic storm). Classes are based on  $Dst$  and differential  $Dst$  time series. Table 1 lists the  $Dst$  class ranges used in this study, the total number of samples per class, and the numbers used for testing and training.

**Table 1.**  $Dst$ -based classification rules used in this study, the total number of samples in each class, and the number of samples used for testing and training. The upper limit is excluded, while the lower limit is included.

| $Dst$        | Storm phase  | Total   | Test    | Train   |
|--------------|--------------|---------|---------|---------|
| $\geq 15$    | positive (P) | 8,360   | 1,672   | 6,688   |
| $[-20, 15>$  | normal (N)   | 379,167 | 75,832  | 303,335 |
| $[-55, -20>$ | recovery (R) | 115,856 | 23,168  | 92,688  |
| $[-85, -55>$ | through (T)  | 5,666   | 1,132   | 4,534   |
| $< -85$      | extreme (E)  | 832     | 164     | 668     |
| <b>Any</b>   | Any          | 509,881 | 101,968 | 407,913 |

To address the differing temporal resolutions of the input variables, such as hourly *Dst*, daily *A*-index, and 30-second RINEX data, all datasets were synchronised to the cadence of the high-frequency RINEX TEC measurements. The *Dst* index values and the daily *A*-index can be repeated as needed.

522,297 represents the total count of TEC observations available in the dataset. Of 522,297 original samples, 12,416 were identified as outliers in exploratory statistical analysis, as their *TEC* is 300 TECU or more. *TEC* values exceeding 300 TECU are usually due to measurement errors or errors in the *TEC* estimation process. 509,881 samples are used for training and testing, with an 80:20 split. The division was stratified using the "createDataPartition" function from the R library [20] "caret" [19]. The training dataset was further divided into training and validation sets during 10-fold cross-validation, with 10 repeats, for all models, using the "trainControl" function from the "caret" package [19]. Samples were split into 5 class ranges, based on *Dst* values derived from theoretical knowledge of different storm phases. A similar class division was proposed by Loewe and Pröls [28]. It is evident from Table 1 that the normal (N) and recovery (R) classes are more common than those with very high or very low *Dst* values, which impacts model performance. The values "scale" and "centre" were used in this study in the "preProcess" parameter for the train function from the R package [20] "caret" [19]. The option "center" subtracts the mean of each feature, while "scale" divides by the standard deviation.

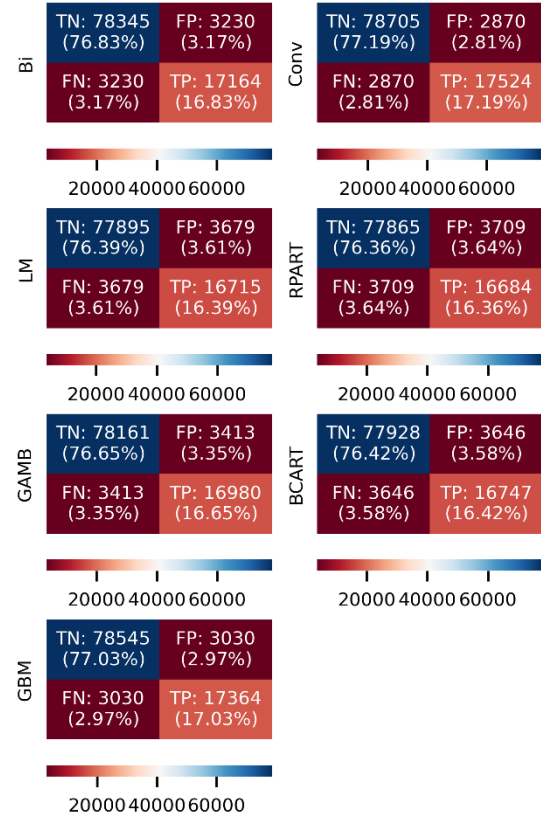
#### 4. Results

Candidate models were evaluated to determine the optimal method and set of predictors to be included in the final model for application in real-world settings. Research results are presented with the following set of six predictors: the Total Electron Content (*TEC*), the standard deviation of Total Electron Content (*dTEC*),  $B_x$ ,  $B_y$ ,  $B_z$ , and  $a_p$ . Theoretical studies show that geomagnetic indices  $B_x$ ,  $B_y$ , and  $B_z$  affect *Dst*, and, in turn, *TEC*. This is why geomagnetic indices ( $B_x$ ,  $B_y$ , and  $B_z$ ), *TEC*, *dTEC*, and  $a_p$  were used as predictors in the final model.

The evaluated methods using the full set of predictors are abbreviated as follows in further text: the LSTM deep learning method with bidirectional layers (Bi), the LSTM deep learning method with convolutional layers (Conv), the linear model (LM), the classification and regression trees implemented in the "rpart" package (RPART), the generalised additive model boosting (GAMB) method, the bagged classification and regression trees (BCART), and the gradient boosting machine (GBM) method.

For multiple classes, the values for each metric were calculated by a "one versus all" approach, viewing each

class as positive and all others as the opposite, and then producing the arithmetic average across all classes, presented in Figure 3 as a bar chart and in Figure 2 as a confusion matrix heatmap.



**Fig. 2.** The confusion matrix heatmap with the number and percentage (%) of TN, FP, FN, and TP classifications for the arithmetic average of all the *Dst*-index classes.

The Conv model had the highest numbers and percentages (%) of TP and TN classifications and the lowest numbers and percentages (%) of FP and FN classifications, while the opposite is true for the RPART model. The Conv model had the highest TPR, TNR, PPV, NPV, accuracy, BA, and *F1* score, while the RPART model had the worst performance across the *Dst*-index classes and is thus unsuitable for this problem. The arithmetic average across all *Dst*-index classes clearly positions the Conv model as the most accurate, with the highest scores across all metrics. In contrast, the RPART model achieved the lowest scores across all metrics, confirming that it cannot be used for the application intended in this study.



**Fig. 3.** The bar charts of the TPR, TNR, PPV, NPV, Acc, BA, and F1 score for the arithmetic average of all the Dst-index classes. The 95% CI is in square brackets. The highest scores for each metric are marked in bold and green, and the lowest scores are bold and red.

The interpretation of the results must also weigh statistical rigour against computational feasibility. The total training and testing time for each candidate model is presented in Table 2. The total model training and testing time, in seconds (s), for all samples, utilising the R "system.time" function [20], across all candidate models, is displayed in Table 2.

The LM method required the least testing time because of its simplicity, and the BCART method required the least

training time. The Bi and Conv methods have the highest computational demands due to deep learning implementation. The Conv model required 61.31 s for training (on the test setup) compared to just 1.22 s for BCART. While the Conv model is computationally intensive, the faster models (LM and RPART) are disqualified because they cannot detect the safety-critical extreme phases of ionospheric storms.

**Table 2.** The total training and testing time for all methods, with the lowest values marked in bold.

| Time  | Bi      | Conv    | LM            | RPART   | GAMB    | BCART         | GBM    |
|-------|---------|---------|---------------|---------|---------|---------------|--------|
| Train | 35.68 s | 61.31 s | 3.8 s         | 12.31 s | 36.44 s | <b>1.22 s</b> | 1.39 s |
| Test  | 2.45 s  | 2.23 s  | <b>0.03 s</b> | 0.77 s  | 7.14 s  | 1.66 s        | 1.32 s |

## 5. Conclusion

This work introduced and validated a novel machine learning framework for classifying geomagnetic storm phases using the *Dst*-index and related ionospheric and geomagnetic variables, with the overarching goal of supporting GNSS PNT integrity in the face of space weather disturbances. This work contributes toward enabling (AA)<sup>2</sup> PNT systems through models deployable on smartphone-class platforms. This study compares traditional statistical models, tree-based ensembles, and deep learning networks. Five discrete *Dst*-based storm phases are identified and used as labels consistent with prior work. The findings show that deep learning architectures, such as convolutional and bidirectional LSTM models, outperform the competition, particularly for identifying rare and extreme space weather events that are critical to PNT reliability. These models successfully captured complex, non-linear relationships between GNSS-relevant features (*TEC*, *dTEC*, *B<sub>x</sub>*, *B<sub>y</sub>*, *B<sub>z</sub>*, and *a<sub>p</sub>*) and *Dst*-index storm classifications, outperforming simpler methods that failed to generalise to minority classes. The accuracy benefits come with the caveat of a higher computational cost. The presented research offers guidelines for selecting appropriate models based on deployment context, including cloud-based systems and embedded platforms. This work builds on previous contributions that warned of the need for a local, context-dependent GNSS model. The developed methodology for classifying space weather events represents a foundation for novel ambient-aware correction approaches to improve GNSS performance in the next generation of systems. The use extends to aviation, autonomous vehicles, maritime navigation, and emergency response. Future research directions include expanding the dataset to include more locations and dates, implementing more complex resampling methods, applying anomaly detection, or adopting hybrid ensemble strategies.

## References

- [1] Parkinson, B.W. & Spilker Jr., J.J. (1996). *Global Positioning System: Theory and Applications, Volume I*. American Institute of Aeronautics and Astronautics, Reston, VA, USA.
- [2] Filjar, R., Damas, M.C., & Iliev, T.B. (2020). Resilient Satellite Navigation Empowers Modern Science, Economy, and Society, *CIEES 2020*, Borovets, Bulgaria, 2020, *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1032, article no. 012001, doi: 10.1088/1757-899X/1032/1/012001.
- [3] Schaer, S. (1999). *Mapping and predicting the Earth's ionosphere using the Global Positioning System*. Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule, Zürich, Switzerland.
- [4] Oxley, A. (2017). *Uncertainties in GPS Positioning: A Mathematical Discourse*. Academic Press/Elsevier, London, UK.
- [5] El Yousoufi, H., Emran, A. & Filjar, R. (2024). Identification of GPS PNT degradation using analysis of non-linear properties of GPS-derived TEC time series, *ION ITM 2024*, Long Beach, CA, pp. 135-145, doi: 10.33012/2024.19527.
- [6] Filjar, R. (2022). An application-centred resilient GNSS position estimation algorithm based on positioning environment conditions awareness, *ION ITM 2022*, Long Beach, CA, pp. 1123-1136, doi: 10.33012/2022.18247.
- [7] Davies, K. (1990). *Ionospheric Radio*. Peter Peregrinus Ltd., London, UK.
- [8] Klobuchar, J.A. (1987). Ionospheric time-delay algorithm for single-frequency GPS users, *IEEE Trans. on Aerosp. and Electronic Systems*, vol. 23, pp. 325-331, doi: 10.1109/TAES.1987.310829.
- [9] Žužić, L., Škrlić, L., Nešković, A. & Filjar R. (2025). Characterisation of GPS Horizontal Positioning Errors and *Dst* Using Recurrence Plot Analysis in Sub-Equatorial Ionospheric Conditions, *Urban Sci.*, vol. 9, no. 11, article no. 451, doi: 10.3390/urbansci9110451.
- [10] Filić, M. & Filjar, R. (2018). Modelling the Relation between GNSS Positioning Performance Degradation, and Space Weather and Ionospheric Conditions using RReliefF Features Selection, *31st International Technical Meeting ION GNSS+ 2018*, Miami, FL, pp. 1999-2006, doi: 10.33012/2018.16016.
- [11] Filjar, R. (2022). A contribution to short-term rapidly developing geomagnetic storm classification for GNSS ionospheric effects mitigation model development, *ICEASE 2021 Conference*, Islamabad, Pakistan, doi: 10.1109/ICASE54940.2021.9904168.
- [12] Filjar, R. (2022). On performance assessment of machine learning-based GNSS ionospheric delay correction model, *International Workshop on Machine Learning for Space Weather: Fundamentals, Tools, and Future Prospects*. International Centre for Theoretical Physics (ICTP), Buenos Aires, Argentina.
- [13] Seemala, G. (2024). GPS-TEC analysis program (version 3.5). <https://seemala.blogspot.com/> (accessed on 3 September 2025).
- [14] INTERMAGNET. (2024). International Real-time Magnetic Observatory Network. <https://intermagnet.org/> (accessed on 3 September 2025).
- [15] Kuhn, M., Johnson, K. (2013). *Applied Predictive Modeling*. Springer Nature, Chamomix, Switzerland.
- [16] Biecek, P. & Burzykowski, T. (2021). *Explanatory Model Analysis*. CRC Press, Boca Raton, FL.
- [17] Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press, Cambridge, UK.
- [18] Boehmke, B. & Greenwell, B. (2020). *Hands-On Machine Learning with R*. CRC Press, Boca Raton, FL.
- [19] Kuhn, M. (2007). caret: Classification and Regression Training, doi: 10.32614/cran.package.caret (accessed on 3 September 2025).
- [20] R Core Team (2025). R: A Language and Environment for Statistical Computing, version 4.5.1. <https://www.r-project.org> (accessed on 3 September 2025).
- [21] Njirjak, M., Žužić, L., Babić, M., Janković, P., Otović, E., Kalafatović, D. & Mauša, G. (2024). Reshaping the discovery of self-assembling peptides with generative AI guided by hybrid deep learning, *Nat. Mach. Intell.*, vol. 6, pp. 1487-1500, doi: 10.1038/s42256-024-00928-1.
- [22] Fennell, P.G., Zuo, Z. & Lerman, K. (2019). Predicting and explaining behavioral data with structured feature space decomposition, *EPJ Data Science*, vol. 8, pp. 2-3, doi: 10.1140/epjds/s13688-019-0201-0.
- [23] Singh, D. & Singh, B. (2020). Investigating the impact of data normalization on classification performance, *Appl. Soft Computing*, vol. 97, article no. 105524, doi: 10.1016/j.asoc.2019.105524.
- [24] Otović, E., Njirjak, M., Kalafatović, D. & Mauša, G. (2022). Sequential Properties Representation Scheme for Recurrent Neural Network-Based Prediction of Therapeutic Peptides, *J. of Chemical Inf. and Modeling*, vol. 62, pp. 2961-2972, doi: 10.1021/acs.jcim.2c00526.
- [25] Altché, F. & de La Fortelle, A. (2017). An LSTM network for highway trajectory prediction, *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*,

- Yokohama, Japan, pp. 1929-1958, doi: 10.5555/2627435.2670313.
- [26] Matzka, J., Bronkalla, O., Tornow, K., Elger, K. & Stolle, C. (2021). Geomagnetic Kp index, doi: 10.5880/Kp.0001 (accessed on 3 September 2025).
- [27] Zolesi, B. & Cander, Lj. R. (2014). Ionospheric prediction and forecasting, Springer Nature, London, England.
- [28] Loewe, C. A. & Prölss, G. W. (1997). Classification and mean behavior of magnetic storms, *Journal of Geophysical Research: Space Physics*, vol. 102, no. A7, pp. 14209-14213, doi: 10.1029/96JA04020.
- [29] Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T. & Vasyliunas, V. M. (1994). What is a geomagnetic storm?, *Journal of Geophysical Research: Space Physics*, vol. 99, no. A4, pp. 5771-5792, doi: 10.1029/93JA02867.
- [30] Kamide, Y., Yokoyama, N., Gonzalez, W., Tsurutani, B. T., Daglis, I. A., Brekke, A. & Masuda, S. (1998). Two-step development of geomagnetic storms, *Journal of Geophysical Research: Space Physics*, vol. 103, no. A4, pp. 6917-6921, doi: 10.1029/97JA03337.
- [31] Rozhnoi, A., Solovieva, M. S., Molchanov, O. A. & Hayakawa, M. (2004). Middle latitude LF (40 kHz) phase variations associated with earthquakes for quiet and disturbed geomagnetic conditions, *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 29, no. 4-9, pp. 589-598, doi: 10.1016/j.pce.2003.08.061.
- [32] Contadakis, M. E., Arabelos, D. N., Pikridas, C. & Spatalas, S. D. (2012). Total electron content variations over southern Europe before and during the M 6.3 Abruzzo earthquake of April 6, 2009, *Annals of geophysics*, vol. 55, no. 1, doi: 10.4401/ag-5322.