

Automated Ionospheric Anomaly Detection During a Geomagnetic Storm Using a Hybrid Wavelet Isolation Forest Approach with Explainable Artificial Intelligence

Syarifah Noor Syakiylla Sayed Daud¹, Kamarul Hawari Ghazali²

¹Centre for Advanced Industrial Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan
Pahang Darul Makmur, Malaysia

²Faculty of Electrical and Electronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan
Pahang Darul Makmur, Malaysia

Article Info

Article history:

Received July 29, 2026

Revised September 15, 2026

Accepted September 20, 2026

Keywords:

Artificial intelligence

Discrete wavelet transforms

Ionosphere anomaly

Feature importance

Geomagnetic storm

ABSTRACT

Detection of anomalies in the ionosphere is essential for ensuring reliable timing, navigation, and positioning services of the Global Navigation Satellite System (GNSS). This study proposes a machine learning approach to anomaly detection during severe geomagnetic storms based on the Wavelet Transform with the Isolation Forest algorithm. Explainable Artificial Intelligence (XAI) methods, specifically Shapley Additive exPlanations (SHAP), are employed to interpret feature contributions and improve model transparency. Total Electron Content (TEC) time-series data were derived from the ionospheric dataset of the geomagnetic storm event for time-frequency analysis. Then, an Isolation Forest model is applied to identify anomalous patterns based on mean, variance, standard deviation, kurtosis, entropy, and energy to represent the temporal dynamics of the ionosphere. Results showed that the model identified 3% of anomalies in the ionosphere with the highest detection accuracy of 96%. Among the wavelet features, the entropy and energy are found to be the most important for anomaly detection. Overall, the proposed model provides a reliable and interpretable solution for unsupervised ionospheric anomaly detection, with potential applications in GNSS monitoring and space weather research.

Corresponding Author:

Kamarul Hawari Ghazali

Faculty of Electrical and Electronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan Pahang Darul Makmur, Malaysia

Email: kamarul@ump.edu.my

1. INTRODUCTION

The ionosphere is a dynamic region of the Earth's upper atmosphere that contains a high concentration of free electrons and ionized particles produced by solar radiation. Fluctuations and rapid changes in ionospheric electron density significantly affect the propagation of GNSS signals [1], [2]. This causes positioning degradation, signal delays, and communication instability [3], [4]. Specifically, a geomagnetic storm is a temporary disturbance of the Earth's magnetic field triggered by enhanced solar activity, such as coronal mass ejections (CMEs) or high-speed solar wind streams. These disturbances occur when charged solar particles interact with the Earth's magnetosphere, causing variations in the geomagnetic field. Because GNSS signals propagate through the ionosphere, changes in electron density directly impact signal transmission characteristics.

TEC is commonly used to monitor ionospheric behaviour, as it reflects the total electron concentration encountered by GNSS signals during propagation [5], [6]. TEC measurements enable monitoring of electron density variations and assessing ionospheric disturbances. Under quiet geomagnetic conditions, TEC generally exhibits smooth temporal variation. In contrast, disturbed conditions can produce abrupt TEC fluctuations, irregular plasma structures, and increased signal variability, driven by heightened geomagnetic and solar activity. Detecting such ionospheric anomalies is essential for understanding space weather phenomena and improving GNSS reliability. The geomagnetic storm on 17 March 2015, known as the St. Patrick's Day Storm, was among the most intense geomagnetic disturbances recorded during Solar Cycle 24 [7], [8], [9]. This storm was primarily triggered by a CME emitted from the Sun on 15 March 2015, which interacted with the Earth's magnetosphere approximately two days later.

Traditional ionospheric anomaly detection methods primarily use statistical analysis and time-series modelling, including Z-score techniques, sliding interquartile range, moving-average residuals, rate of TEC index, and empirical threshold models [10], [11]. While these approaches are computationally efficient and well-established, they frequently fail to capture the complex, non-stationary, and multi-scale characteristics of ionospheric signals, particularly during intense geomagnetic disturbances. Methods that operate exclusively in the time domain are often inadequate for detecting localized transient variations. Consequently, wavelet-based techniques have been increasingly adopted to enhance the detection and characterization of ionospheric anomalies. Furthermore, recent advancements in machine learning have attracted significant attention for anomaly detection, owing to their ability to uncover hidden patterns and nonlinear relationships within large datasets. Among these, unsupervised learning methods are especially promising, as ionospheric datasets typically lack reliable ground-truth labels.

Wavelet Transform is widely recognized as an effective signal processing technique for analysing non-stationary signals, owing to its multi-resolution capability in both time and frequency domains [12], [13], [14]. In ionospheric research, the Wavelet Transform is frequently employed for TEC characterization, feature extraction, and denoising. High-frequency wavelet detail coefficients are particularly valuable for detecting localized TEC perturbations generated during geomagnetic storm events, traveling ionospheric disturbances (TIDs), plasma irregularities, and pre-seismic ionospheric anomalies. Statistical features derived from wavelet coefficients provide meaningful representations of ionospheric behaviour, which are beneficial for machine learning applications.

Kumar Dabbakuti and Venkata Ratnam [15] identified ionospheric TEC variability using wavelet transform analysis and reported that wavelet methods effectively characterize disturbed ionospheric behaviour under varying geomagnetic conditions. Similarly, Krankowski et al. [16] utilized wavelet analysis to identify disturbed TEC conditions across multiple GPS stations. They demonstrated the capability of wavelet techniques to localize temporal TEC variations and ionospheric disturbances. Mandrikova et al. [17] integrated wavelet transforms with neural network approaches for ionospheric disturbance analysis and anomaly detection, indicating that wavelet-based preprocessing enhances the identification of disturbed ionospheric periods. Yang et al. [18] also combined wavelet transform with neural network approaches for ionospheric disturbance analysis and anomaly detection, confirming that wavelet-based preprocessing improves the identification of disturbed ionospheric periods. Additionally, Mandrikova et al. [19] employed wavelet transform in conjunction with autoregressive modelling to identify anomalous ionospheric behaviour associated with geophysical disturbances.

Isolation Forest is an efficient unsupervised anomaly detection algorithm that identifies anomalies based on the principle of data isolation. The application of this method in ionospheric research is relatively novel and requires further investigation. Isolation Forest has been widely adopted in core industries such as cybersecurity, manufacturing quality control, financial fraud detection, the Internet of Things, and system health monitoring [20], [21], [22], [23]. Its popularity is attributed to linear time complexity and low memory requirements, which enable rapid processing of massive datasets and high-dimensional spaces. Unlike distance-based or density-based methods, Isolation Forest isolates anomalous samples via random partitioning, making it well-suited to detecting rare and irregular ionospheric disturbances without labelled training data. Most previous machine learning-based ionospheric anomaly detection studies have primarily focused on improving detection accuracy, often without providing sufficient interpretability regarding the model's decision-making process. The emergence of XAI techniques presents an opportunity to enhance transparency and trustworthiness in machine learning applications. SHAP is among the most widely adopted XAI approaches for quantifying feature contributions and interpreting model predictions in a mathematically consistent manner [24], [25], [26].

Although wavelet analysis has been widely applied in ionospheric research, several limitations persist in the current literature. Most prior studies have concentrated on TEC characterization, signal denoising, forecasting, or time-frequency visualization, rather than on automated anomaly detection using unsupervised machine learning methods. Additionally, many wavelet-based machine learning models lack interpretability and do not adequately explain the contribution of extracted features to anomaly classification. Existing research

often employs wavelet techniques primarily for visualization, without fully leveraging the statistical information contained in multi-resolution wavelet coefficients. Moreover, the integration of wavelet-derived statistical features with unsupervised anomaly isolation methods remains limited, especially in the context of geomagnetic storm-related ionospheric disturbances. These gaps underscore the need for an explainable, wavelet-based, unsupervised anomaly detection framework, as proposed in this study.

2. MATERIALS AND METHODS

2.1. Dataset description

This study utilizes the public ionospheric TEC dataset obtained during the St. Patrick's Day Storm geomagnetic storm event on 17 March 2015, retrieved from <https://doi.org/10.18710/RSBWOH> [27]. This day is widely recognized as one of the most intense geomagnetic disturbances of Solar Cycle 24. The TEC monitor model GSV 4004 was originally installed at Blindern Campus and was moved to Skibotn in September 2014. The TEC data are calibrated using the GPS-SCINDA software and have a 60 s resolution [28], [29]. Before the analysis of this work, the TEC dataset underwent preprocessing, including the removal of unwanted and missing values to ensure data consistency. Invalid and non-numeric observations were excluded from the analysis to improve signal quality and model reliability.

2.2. Proposed framework for anomaly detection

The proposed framework consists of six major stages, which are TEC data preprocessing, DWT decomposition, statistical feature extraction from wavelet coefficients, feature normalization, anomaly detection using Isolation Forest, and SHAP XAI analysis to evaluate feature contributions. In addition, the conventional Z-score statistical method was adopted as a baseline approach for comparative performance evaluation. The data processing and analysis were implemented using the Python programming language in the Spyder Integrated Development Environment (IDE). The overall workflow of the proposed methodology is illustrated in Figure 1.

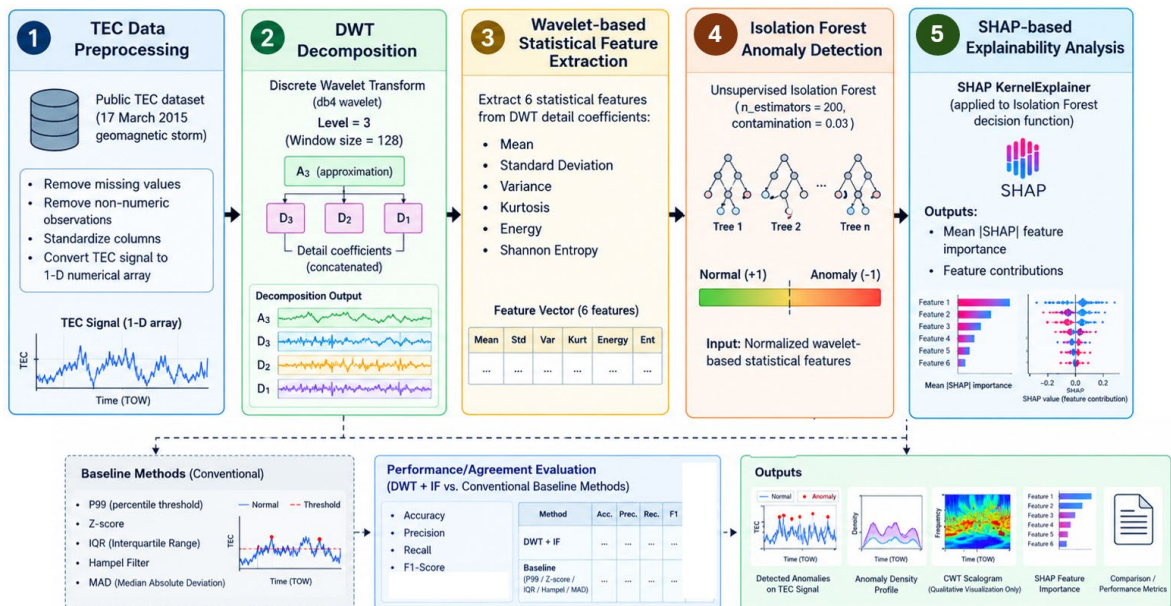


Figure 1. Illustrates the overall workflow of the proposed framework

2.3. Data preprocessing

The TEC dataset was imported into the Python environment using the Pandas library. The parameter of TEC adopted in this work was the TEC at the TOW signal, which represents the temporal variation of ionospheric electron content observed from GNSS measurements. Besides, the Scikit-learn toolkit of NumPy, SciPy, and Matplotlib was also utilized for multidimensional array operations, scientific and technical computing, data visualization, and plotting. The first step taken was to preprocess the imported dataset to provide consistent and improved data quality. The column names of the dataset were standardized by removing unnecessary spacing, whereas missing values and non-numeric observations were removed from the TEC signal dataset. After preprocessing, the TEC signal was converted into a one-dimensional numerical array for subsequent signal decomposition and anomaly analysis.

2.4. Discrete wavelet transform (DWT) decomposition

The original TEC signal was decomposed using the DWT method to determine its multi-resolution characteristics in both time and frequency domains. DWT decomposes the original signal into approximation and detail coefficients at multiple resolution levels. Therefore, it permits the separation of low-frequency background trends from high-frequency disturbance components. The wavelet parameters were selected based on preliminary evaluation, the characteristics of the GNSS TEC signal, and consistency with previous studies. The preliminary evaluation in Table 1 showed that the tested configurations produced relatively comparable anomaly detection results, with no substantial differences in the detected anomaly proportions. Therefore, the parameter selection was not based solely on the anomaly detection output, but also considered the suitability of each parameter for representing the non-stationary and localized variations of GNSS TEC signals. A window size of 128 samples was selected to provide a balance between temporal context and the localisation of transient ionospheric variations. The db4 wavelet was selected because its compact support and shape are suitable for representing localized signal variations, while its use is also consistent with previous wavelet-based GNSS and ionospheric signal studies. A three-level decomposition was selected to provide sufficient multiscale representation without excessive decomposition. Based on these considerations, the 128-sample window, db4 wavelet, and three-level decomposition were adopted for the subsequent analysis.

Table 1. Summary of preliminary wavelet parameter evaluation

Parameter	Tested Values	Selected	Preliminary Result	Anomaly Detection Result
Window size	64, 128, 256	128	Comparable anomaly detection (3.08–4.17%)	128 provided an anomaly proportion closest to the target 3% while retaining sufficient temporal windows
Wavelet	db2, db4, db6	db4	Comparable anomaly detection (3.45–3.63%)	db4 provided a controlled anomaly proportion with stable detection across decomposition levels
Decomposition level	2, 3, 4	3	Comparable anomaly detection (3.45–3.63%)	Level 3 provided a balanced multiscale representation without increasing the detected anomaly proportion substantially

In this study, the Daubechies wavelet of order four (db4) was selected due to its suitability and effectiveness in representing transient signal variations and its widespread application in geophysical signal analysis [30]. The TEC signal was decomposed up to level 3 using the wavelet decomposition process, where one approximation coefficient (A_3) and three detail coefficients (D_3 , D_2 , and D_1) allow the signal to be analysed across multiple frequency scales, as in (1):

$$x(t) = A_3 + D_3 + D_2 + D_1 \quad (1)$$

where A_3 represents the approximation coefficient at level 3, while D_1 , D_2 , and D_3 denote detail coefficients corresponding to different frequency bands.

A sliding window segmentation technique with a window size of 128 samples was subsequently applied to extract localized TEC variations throughout the observation period. The detailed coefficients from all decomposition levels were concatenated and utilized for statistical feature extraction. The overall workflow using the DWT method is represented in Figure 2.

Figure 3 shows the TEC signal after being decomposed into approximation and detail coefficients. The approximation coefficient (A_3) indicates the low-frequency component of the TEC signal with smoother properties. The output revealed that the Discrete Wavelet Transform successfully decomposed the TEC signal into approximation and detail coefficients, allowing the separation of large-scale background trends from localized high-frequency disturbances. The extracted detail coefficients contain rapid temporal fluctuations that are strongly related to ionospheric irregularities and storm-induced perturbations.

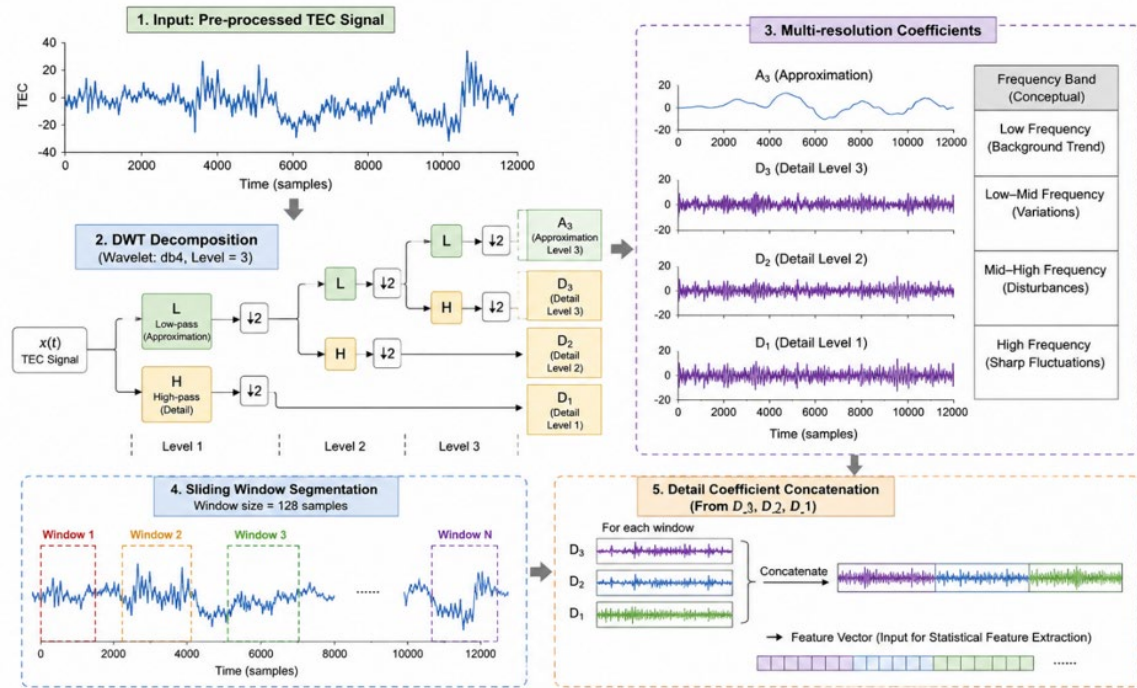


Figure 2. Workflow of the DWT decomposition process

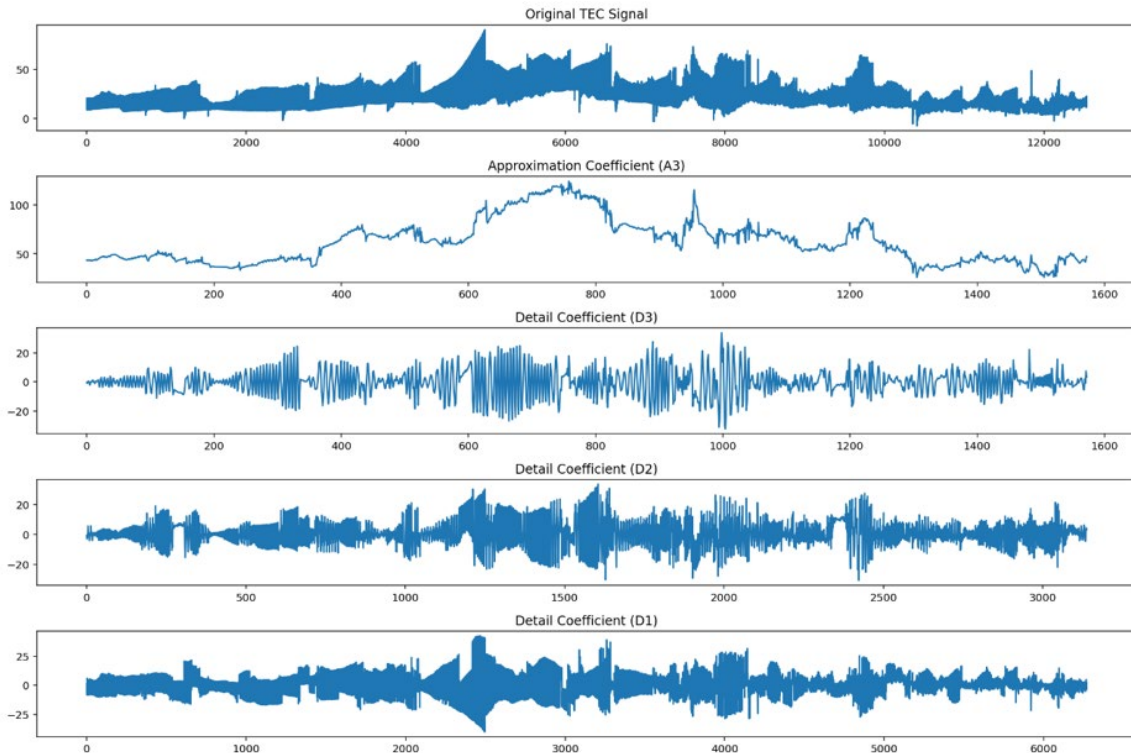


Figure 3. Decomposition of TEC signal into approximation and detail coefficients using the DWT method

2.5. Wavelet-based statistical feature extraction

A few selected statistical features were extracted from the DWT detail coefficients to assess the ionospheric signal behaviour. The obtained features include mean, standard deviation, variance, kurtosis, energy, and Shannon entropy. These features will provide information on the intensity, dispersion, impulsiveness, and complexity of ionospheric disturbances during geomagnetic storms.

The mean was used to represent the average value of the wavelet coefficients and is denoted as in (2):

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

where x_i denotes the individual wavelet coefficients and N is the total number of coefficients.

After that, the standard deviation is also obtained to quantify the spread of the wavelet coefficients around the mean and is defined as in (3):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (3)$$

where σ represents the standard deviation of the wavelet coefficients.

Then, variance was determined to measure the dispersion of wavelet coefficients relative to the mean value and is expressed as in (4):

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (4)$$

where x_i represents the wavelet coefficient, μ is the mean value, and N denotes the total number of coefficients. Kurtosis is also extracted to measure the impulsiveness and the sharpness of the signal distribution, and is denoted as in (5):

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 \quad (5)$$

where x_i represents the wavelet coefficients, μ is the mean, and σ denotes the standard deviation. High kurtosis values indicate sharp, transient fluctuations and abnormal signal behaviour.

The wavelet energy was calculated as in (6):

$$E = \sum_{i=1}^N D_i^2 \quad (6)$$

where D_i represents the wavelet detail coefficients. Shannon entropy was used to measure signal randomness and complexity, represented as in (7):

$$H = - \sum_{i=1}^N p_i \log(p_i) \quad (7)$$

where p_i denotes the probability distribution of wavelet coefficients. The extracted features were organized into a feature matrix and normalized using StandardScaler in Python to ensure consistent feature scaling before machine learning analysis.

2.6. Feature normalization

A StandardScaler method from the Scikit-learn library of Python was used to normalize the extracted statistical feature matrix to allow consistent feature scaling before machine learning analysis. Feature normalization standardizes the feature distribution to zero mean and unit variance, thereby preventing features with larger numerical ranges from dominating the anomaly detection process. The normalization process can be expressed as in (8):

$$z = \frac{x - \mu}{\sigma} \quad (8)$$

where x represents the original feature value, μ is the feature mean, and σ is the standard deviation.

2.7. Isolation forest for anomaly detection

This study used Isolation Forest as the unsupervised anomaly detection model. Unlike conventional classification approaches, Isolation Forest identifies anomalies by isolating observations through random partitioning. Anomalous samples are generally isolated faster than normal samples due to their rarity and distinct feature distribution. The Isolation Forest model was configured using 200 decision trees with a contamination parameter of 0.03, representing the estimated proportion of anomalies within the dataset. The normalized wavelet-derived statistical features were used as the input for model training and anomaly classification. The prediction output is represented as in (9):

$$f(x) \in \{-1, 1\} \quad (9)$$

where -1 indicates anomalous observations and 1 denotes normal observations. The detected anomalies were subsequently visualized on the TEC signal to facilitate the interpretation and analysis of ionospheric disturbances.

A preliminary evaluation was conducted to determine an appropriate contamination value for the GNSS anomaly detection task. Five contamination values, namely 0.01, 0.02, 0.03, 0.05, and 0.10, were evaluated while maintaining a fixed number of 200 decision trees. The evaluation focused on the resulting number and proportion of detected anomalies to identify a setting that could capture rare observations without producing excessive anomaly detections. As shown in Table 2, the detected anomaly proportion increased progressively from 1.01% to 10.00% as the contamination value increased from 0.01 to 0.10.

Table 2. Evaluation of various statistical anomaly detection models in comparison with the proposed IF model

Anomaly Detection Model	No. of Detected Anomalies	Percentage Anomalies
Baseline Model		
Median Absolute Deviation (MAD)	532	4.29%
Interquartile Range (IQR)	452	3.64%
Hampel Filter	562	4.52%
99 th Percentile (P99)	126	1.02%
z-score	243	0.96%
Proposed Model		
IF with contamination 0.01	125	1.01%
IF with contamination 0.02	249	2.01%
IF with contamination 0.03	373	3.01%
IF with contamination 0.05	621	5.00%
IF with contamination 0.10	1241	10.00%

A contamination value of 0.03 was selected because it resulted in 373 detected anomalies, corresponding to 3.01% of the observations. This anomaly proportion was considered representative of the expected rarity of anomalous GNSS observations and was also within the range produced by the statistical baseline methods. In particular, the Median Absolute Deviation (MAD), Interquartile Range (IQR), and Hampel Filter identified 4.29%, 3.64%, and 4.52% anomalies, respectively. In contrast, contamination values of 0.05 and 0.10 resulted in substantially higher anomaly proportions of 5.00% and 10.00%, respectively, which may lead to excessive anomaly detection. The 0.01 setting, although comparable to the P99 and z-score baselines, resulted in a relatively small anomaly proportion of 1.01%. Therefore, a contamination value of 0.03 provided a reasonable balance between conservative detection and excessive anomaly identification and was adopted for the subsequent Isolation Forest analysis.

2.8 SHAP-based explainability analysis

To improve the interpretability of the proposed model, SHAP was employed to quantify the contribution of each wavelet-derived statistical feature to the anomaly predictions generated by the Isolation Forest model. SHAP is an XAI technique based on cooperative game theory that assigns an importance value to each feature according to its contribution to the model output.

In this study, SHAP values were computed using the KernelExplainer method, which provides model diagnostic explanations for the Isolation Forest decision function. Features with higher absolute SHAP values have a greater influence on the anomaly detection results, whereas features with lower SHAP values contribute less to the model predictions.

By integrating SHAP analysis into the proposed framework, the decision-making process of the Isolation Forest model becomes more transparent and interpretable, enabling the identification of the most influential wavelet-derived statistical features associated with ionospheric disturbances.

3. RESULTS AND DISCUSSION

3.1. Pattern of TEC signal during geomagnetic storm

Figure 4 shows the temporal variation of the TEC signal, with anomalies identified by the proposed DWT and Isolation Forest models, along with their corresponding anomaly densities. The detected anomalies are not randomly distributed throughout the observation period but instead exhibit clear temporal clustering. The largest concentration of anomalies occurs in the middle section of the TEC record, where the signal increases

rapidly, followed by pronounced fluctuations. This behaviour indicates a period of substantial ionospheric disturbance characterized by rapid variations in electron density rather than gradual background changes.

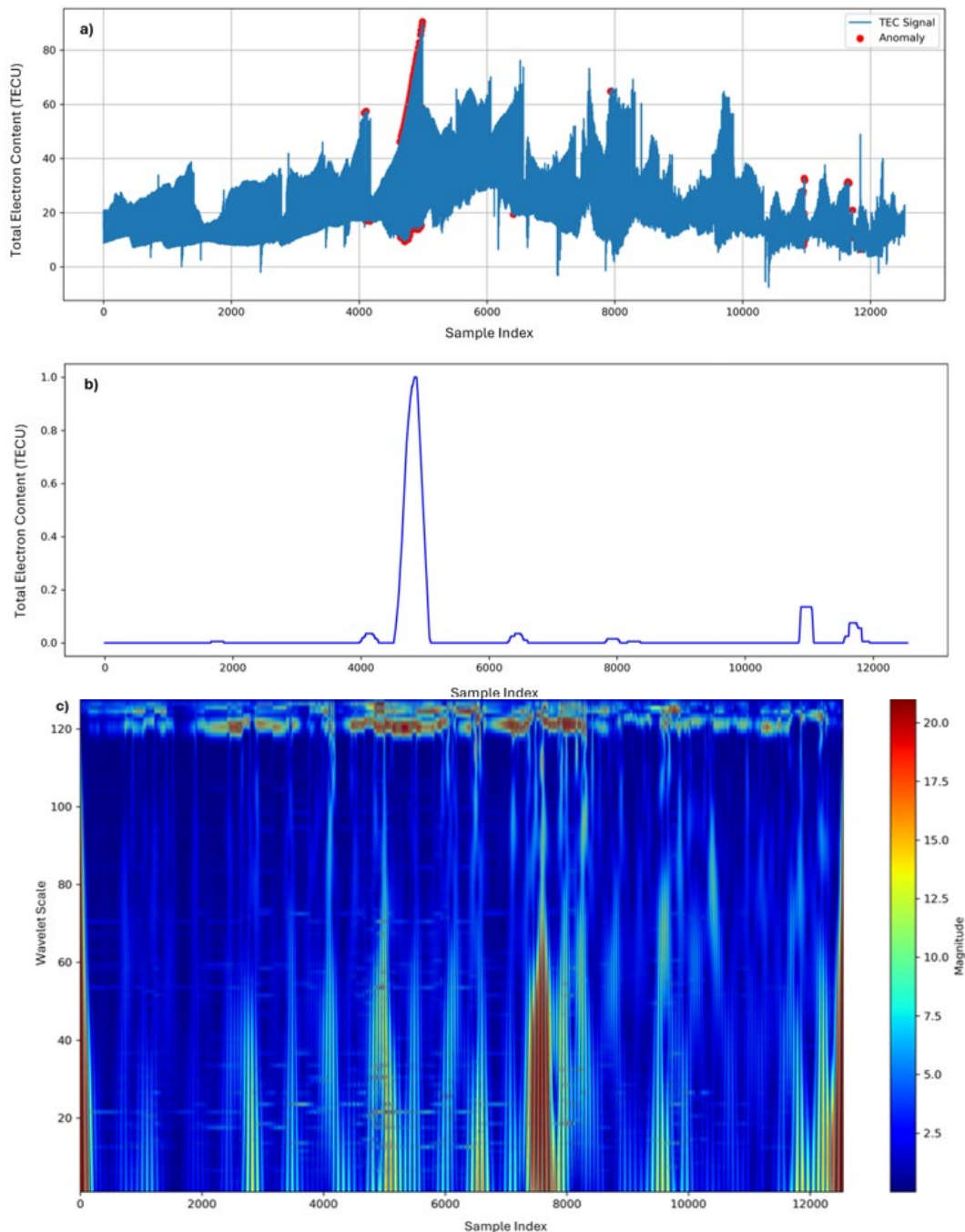


Figure 4. Model output for representation of: (a) Original TEC signal with detected anomalies (red markers), (b) anomaly density profile generated from the detected anomalies in TEC using a moving-average window, and (c) Continuous Wavelet Transform (CWT) scalogram of the TEC signal, presented for qualitative visualization of the time-frequency characteristics during the geomagnetic storm.

The anomaly density profile further confirms this observation by presenting a distinct dominant peak centred within the same interval. The sharp increase in anomaly density indicates that a large number of consecutive data segments were simultaneously classified as anomalous, suggesting the presence of a sustained disturbance rather than isolated abnormal observations. Outside this interval, the anomaly density remains close to zero, with only several small peaks appearing near the later part of the observation period. These secondary peaks correspond to localized disturbances with considerably lower intensity than the dominant event.

An important observation is that anomaly detection is governed by the local statistical characteristics of the wavelet-derived features rather than solely by the magnitude of the TEC signal. Consequently, not every high TEC value is classified as anomalous, while several anomalies are identified during moderate TEC levels where rapid local fluctuations occur [31], [32]. This behaviour demonstrates that the proposed model is capable of distinguishing abnormal signal dynamics from normal ionospheric variability, thereby reducing the likelihood of falsely identifying naturally elevated TEC values as disturbances.

Although the proposed anomaly detection framework is based on DWT-derived statistical features, the CWT scalogram shown in Figure 4 is presented solely for qualitative visualization of the TEC signal. High-intensity regions observed in the scalogram correspond to localized increases in wavelet magnitude, indicating periods of enhanced ionospheric disturbance.

The scalogram reveals that anomalous TEC behaviour occurs across multiple scales, suggesting the presence of both short-term and medium-scale ionospheric irregularities. These localized high-energy structures are consistent with plasma turbulence and rapid electron density fluctuations commonly observed during severe geomagnetic storm conditions. Compared to the raw TEC signal, the CWT representation provides clearer visualization of transient disturbances and demonstrates the effectiveness of wavelet analysis for investigating non-stationary ionospheric behaviour.

Overall, the anomaly distribution and density indicate that the proposed model effectively identifies both the primary storm-related disturbance and smaller localized ionospheric irregularities throughout the observation period. The combination of wavelet-derived statistical features and Isolation Forest enables the model to capture the non-stationary characteristics of the TEC signal while preserving sensitivity to transient ionospheric disturbances. However, relying visually on the signal alone is unable to comprehensively describe the efficiency of the model in detecting anomalies; therefore, quantitative measurements are required.

3.2. Detection of anomalies during geomagnetic storm based on isolation forest classification

Isolation Forest (IF) detects anomalies based on the principle that unusual observations are easier to isolate than normal observations through random partitioning [33]. The IF model successfully identified 373 anomalous observations, corresponding to 3.01% of the total dataset. This relatively small proportion indicates that most of the TEC/ROTI measurements exhibited normal ionospheric behaviour. In contrast, only a limited number of observations demonstrated abnormal signal characteristics. Such a distribution is expected in ionospheric monitoring applications, where stable conditions dominate the time series. Meanwhile, disturbances occur only during specific periods associated with enhanced ionospheric variability.

In this study, normal ionospheric variations generally follow continuous and smooth temporal patterns, whereas anomalous events are characterised by abrupt changes, irregular fluctuations, or deviations from the typical signal distribution [34], [35]. Therefore, the detected 373 anomalies represent observations that exhibited distinct statistical and temporal behaviours compared with the majority of the dataset. The detection rate of approximately 3% demonstrates that the model maintained a conservative detection behaviour and did not excessively classify normal ionospheric fluctuations as anomalies. This characteristic is important for ionospheric monitoring systems because excessive false alarms may reduce the reliability of the detection framework.

Table 3. Agreement with proposed method

Metrics (%)	P99	z-score	IQR	Hampel	MAD
Accuracy	96.53	95.81	94.54	93.43	93.93
Precision	9.12	12.87	19.84	16.09	20.38
Recall	26.98	19.75	16.37	10.68	14.29
F1-Score	13.63	15.58	17.94	12.83	16.80

Table 3 shows that the proposed DWT + IF method achieved relatively high accuracy of agreement with all conventional baseline methods, ranging from 93.43% to 96.53%. The highest accuracy was obtained for P99 (96.53%), followed by Z-score (95.81%), IQR (94.54%), MAD (93.93%), and Hampel (93.43%). The high accuracy values indicate that the proposed method and the baseline approaches generally agreed on the majority of samples, which is expected because normal TEC observations constitute the dominant portion of the dataset.

However, the relatively low precision and recall values indicate that the agreement was considerably lower when focusing specifically on anomaly samples. Among the baselines, MAD produced the highest precision (20.38%), followed by IQR (19.84%), Hampel (16.09%), Z-score (12.87%), and P99 (9.12%). The higher precision of MAD suggests that a larger proportion of the anomalies identified by DWT + IF were also classified as anomalies by MAD compared with the other baseline methods.

In terms of recall, P99 achieved the highest value at 26.98%, followed by Z-score at 19.75%, IQR at 16.37%, MAD at 14.29%, and Hampel at 10.68%. The higher recall of P99 indicates that P99 captured a larger proportion of the anomaly samples identified by the proposed DWT + IF method. This difference between precision and recall demonstrates that the baseline methods respond differently to TEC variations and therefore identify different subsets of anomalous observations.

Overall, the results demonstrate that no conventional baseline showed strong agreement with DWT + IF at the anomaly level, despite their high overall accuracy. This is particularly evident from the relatively low precision and recall values. The finding suggests that DWT + IF identifies anomaly patterns that are not completely captured by simple statistical threshold-based approaches. This can be attributed to the ability of the DWT-based feature representation to capture localized variations and multi-scale characteristics in the TEC signal, while Isolation Forest detects observations that differ from the dominant feature distribution.

Among the baseline methods, P99 showed the highest overall accuracy and recall, whereas MAD achieved the highest precision. Therefore, P99 appears to provide the closest agreement with DWT + IF in terms of the number of shared anomaly detections, while MAD provides the most precise overlap with the anomalies identified by the proposed method.

3.3. SHAP-based feature importance analysis

To interpret the decision-making behaviour of the Isolation Forest model, SHAP analysis was performed to quantify the contribution of each extracted feature to the anomaly detection process. The mean absolute SHAP values indicate the average impact magnitude of each feature on the model output, where higher values represent greater influence on anomaly identification.

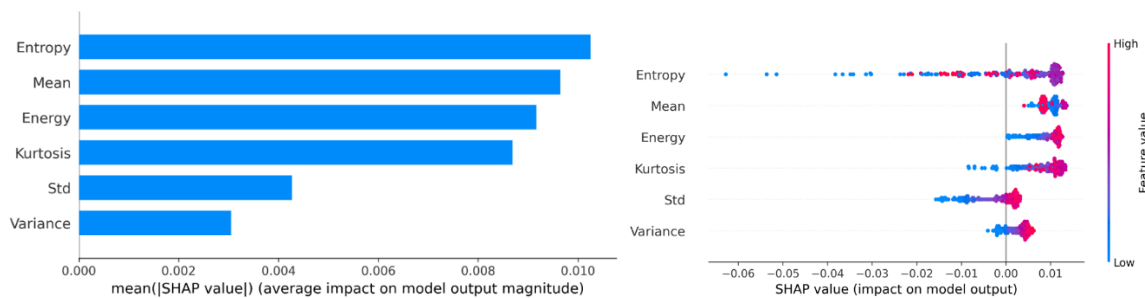


Figure 6. SHAP feature importance ranking showing the contribution of each input feature to the Isolation Forest anomaly detection model.

The SHAP analysis revealed that entropy was the most influential feature, followed by mean, energy, kurtosis, standard deviation, and variance. The dominant contribution of entropy indicates that signal complexity and irregularity play a significant role in distinguishing anomalous ionospheric conditions from normal variations. In wavelet-based analysis, entropy represents the distribution and uncertainty of signal energy across different frequency components. During disturbed ionospheric conditions, TEC signals tend to exhibit increased fluctuations and irregular oscillatory behaviour, resulting in higher complexity that can be effectively captured by entropy.

The mean feature demonstrated the second-highest contribution, suggesting that changes in the average signal level also provide important information for anomaly detection. Significant ionospheric disturbances may produce sustained increases or decreases in TEC values compared with typical background conditions, causing deviations in local signal characteristics. Therefore, mean-based information assists the model in identifying shifts from normal ionospheric behaviour. The energy feature ranked as the third most important variable, highlighting the relevance of signal intensity variations in detecting abnormal events. Wavelet energy reflects the strength of signal fluctuations at different frequency scales, where enhanced disturbance activity may introduce additional signal power associated with ionospheric irregularities.

Meanwhile, kurtosis exhibited a notable contribution, indicating that the distribution shape of the signal contains valuable information for anomaly detection. Higher kurtosis values are associated with impulsive or extreme fluctuations, which may occur during short-duration ionospheric disturbances. In contrast, standard deviation and variance showed relatively lower SHAP contributions, suggesting that overall signal dispersion alone was less effective compared with features describing signal complexity and energy distribution. The dominance of entropy and wavelet-derived energy features confirms the importance of incorporating time-frequency information for characterising complex ionospheric anomalies. These findings support the effectiveness of combining wavelet feature extraction with unsupervised learning, as the model is able to utilise

physically meaningful signal characteristics rather than relying solely on conventional amplitude-based thresholds.

4. DISCUSSION

The present study demonstrates that integrating DWT with Isolation Forest provides an effective approach for automated ionospheric anomaly detection during geomagnetic storm conditions. Unlike conventional statistical approaches that rely on fixed thresholds, the proposed framework analyses wavelet-derived statistical features to characterize localized signal behaviour, allowing the model to identify complex ionospheric disturbances associated with rapid temporal variations in TEC. This capability is particularly important because ionospheric responses to geomagnetic storms are inherently non-stationary and often exhibit multi-scale characteristics that cannot be adequately represented using purely time-domain statistical methods.

The findings are consistent with previous studies that reported the effectiveness of wavelet analysis for representing transient ionospheric disturbances [36]. Earlier works mainly employed wavelet transform for signal decomposition, denoising, or visualization of TEC variations, whereas the present study extends its application by integrating wavelet-derived statistical descriptors into an unsupervised anomaly detection framework. This demonstrates that wavelet features can serve not only as signal representation tools but also as informative inputs for machine learning-based anomaly detection. The incorporation of statistical descriptors such as entropy, energy, kurtosis, and mean enables the model to summarize complex time-frequency information into a compact feature space suitable for automated analysis.

Although the proposed Isolation Forest model achieved high overall accuracy, relatively modest precision and recall values were obtained. This outcome is expected because anomaly detection in ionospheric monitoring represents a highly imbalanced and weakly labelled problem. The reference anomalies generated using the Z-score, Hampel filter, MAD, IQR, and P99 methods do not necessarily represent absolute ground truth but rather provide a statistical baseline for comparison. Consequently, some observations identified by Isolation Forest may correspond to subtle ionospheric disturbances that are not captured by the conventional threshold-based approach. This highlights an inherent challenge in evaluating unsupervised learning algorithms using statistical reference labels, where disagreement between methods does not always indicate incorrect detection.

An important contribution of this study is the incorporation of Explainable Artificial Intelligence through SHAP analysis. Most previous studies focused primarily on improving detection performance without investigating how individual features influence model predictions. The SHAP results demonstrate that entropy and energy contribute more strongly than conventional dispersion-based measures such as variance and standard deviation. This finding suggests that signal complexity and energy redistribution during geomagnetic storms provide more discriminative information than simple amplitude variability. From a physical perspective, disturbed ionospheric conditions introduce irregular plasma dynamics that increase signal randomness and redistribute energy across multiple frequency bands, making entropy and wavelet energy particularly effective indicators of anomalous behaviour.

The proposed framework also offers practical advantages for operational GNSS monitoring. Isolation Forest requires no labelled training data, making it suitable for continuous space weather surveillance where manually annotated anomaly datasets remain scarce. Furthermore, the computational efficiency of both DWT and Isolation Forest allows rapid processing of large TEC datasets, supporting near real-time anomaly detection applications. Such capability may assist GNSS service providers in identifying periods of degraded positioning performance caused by severe ionospheric disturbances.

Despite these encouraging results, several limitations should be acknowledged. The proposed framework was evaluated using a single geomagnetic storm event, which may not fully represent the diversity of ionospheric disturbances occurring under different solar and geomagnetic conditions. In addition, only six wavelet-derived statistical features were investigated. Future studies should validate the proposed framework using multiple storm events collected over different phases of the solar cycle and explore additional wavelet-based descriptors or hybrid deep learning approaches. Comparative evaluation against other unsupervised anomaly detection algorithms, such as One-Class Support Vector Machine and Local Outlier Factor, would also provide a more comprehensive assessment of the proposed methodology.

5. CONCLUSION

This study proposed explainable artificial intelligence approaches for automated ionospheric anomaly detection by integrating Discrete Wavelet Transform, Isolation Forest, and SHAP-based feature interpretation. The proposed approach successfully extracted meaningful wavelet-derived statistical features from TEC signals and identified anomalous ionospheric behaviour associated with the 17 March 2015 geomagnetic storm. Compared with the statistical reference approach, Isolation Forest demonstrated greater flexibility in detecting complex and localized disturbances without relying on predefined statistical thresholds. The SHAP analysis further improved model interpretability by identifying entropy and wavelet energy as the most influential

features contributing to anomaly detection, indicating that signal complexity and energy distribution are more informative than conventional statistical dispersion measures. These findings enhance confidence in the model predictions while providing physical insight into the characteristics of disturbed ionospheric behaviour. Overall, the proposed framework offers a reliable, computationally efficient, and interpretable solution for unsupervised ionospheric anomaly detection. The integration of wavelet-based feature extraction with explainable machine learning provides a promising direction for future GNSS monitoring and space weather applications. Future work will focus on validating the proposed framework using larger multi-year datasets, multiple geomagnetic storm events, and additional unsupervised and deep learning algorithms to further improve robustness and generalization.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Universiti Malaysia Pahang Al-Sultan Abdullah for the financial support provided through the Postdoctoral Fellowship programmed. The author is currently serving as a Postdoctoral Fellow at the university, where the institutional support and research environment have greatly facilitated the progress of this work. This research was conducted under the support of the Strategic Research Fund (SRF) project SRF01251303 funded by Malaysia Ministry of Science, Technology and Innovation (MOSTI), which provided the necessary resources and infrastructure to carry out the study. The author sincerely appreciates the university's commitment to advancing research and innovation.

REFERENCES

- [1] Y. Zeng, D. Zhang, K. Li, Y. Tian and G. Yang, "The Evolution of the Ionospheric Irregularities and Their Effects on GNSS Over China During the Geomagnetic Storm on 1 December 2023," *Space Weather*, vol. 24, no. 4, pp. 1–20, Apr. 2026, doi: 10.1029/2025SW004882
- [2] Y. Liu, K. Yang, L. Sun, J. Wang, A. Smirnov and C. Xiong, "Modeling and Prediction of the Ionosphere with Deep Learning: A Review," *Intelligence and Robotics*, vol. 6, no. 1, pp. 68–96, Mar. 2026, doi: 10.20517/IR.2026.04.
- [3] Y. Kassa, A. Tebabal and B. Damtie, "Nighttime Ionospheric Irregularity During Intense Geomagnetic Storm Events Over The Europe-African Longitudinal Sector," *Heliyon*, vol. 10, no. 19, pp. 1–19, Oct. 2024, doi: 10.1016/J.HELIYON.2024.E38138.
- [4] A. J. Coster, "Ionospheric Irregularities, Space Weather Impact on GNSS, and Spacecraft Anomalies," *Springer Proceedings in Physics*, vol. 431, pp. 131–141, 2026, doi: 10.1007/978-981-95-1121-1_15/FIGURES/2.
- [5] N. Yaacob, M. Abdullah, and M. Ismail, "GPS Total Electron Content (TEC) Prediction at Ionosphere Layer over the Equatorial Region," in *Trends in Telecommunications Technologies*, IntechOpen Limited, London: United Kingdom. 2010, doi: 10.5772/8474.
- [6] S. Shimhanda, S. Francis, and P. J. Cilliers, "Analysis of Ionospheric Total Electron Content Variability Determined from Global Navigation Satellite System Data," *Namibian Journal for Research, Science and Technology*, vol. 1, no. 1, pp. 3–9, Oct. 2020, doi: 10.54421/NJRST.V1I1.4.
- [7] C. C. Wu *et al.*, "The First Super Geomagnetic Storm of Solar Cycle 24: 'The St. Patrick's Day Event (17 March 2015)' Global Data Systems for the Study of Solar-Terrestrial Variability 3. Space Science," *Earth, Planets and Space*, vol. 68, no. 1, pp. 1–12, Sep. 2016, doi: 10.1186/S40623-016-0525-Y.
- [8] Z. Fu, N. Wang, X. Shen and A. Li, "Temporal and Spatial Analysis of the Impact of the 2015 St. Patrick's Day Geomagnetic Storm on Ionospheric TEC Gradients and GNSS Positioning in China Using GIX and ROTI Indices," *Remote Sensing*, vol. 17, no. 12, pp. 1–17, June. 2025, doi: 10.3390/RS17122027.
- [9] N. Mridula, G. Manju, S. Sijikumar, T. Kumar Pant, and R. Kumar Choudhary, "On the Significant Impact of the 17 March 2015 St. Patrick's Day Geomagnetic Storm on the Ionosphere Over Indian Region," *Advances in Space Research*, vol. 70, no. 11, pp. 3674–3685, Dec. 2022, doi: 10.1016/J.ASR.2022.08.022.
- [10] L. Li *et al.*, "Anomalies of Land-Atmospheric Parameters in the 2023 Turkey Earthquake (Mw7.8) using GNSS and ERA5," *Geomatics, Natural Hazards and Risk*, vol. 17, no. 1, pp. 1–26, Feb. 2026, doi: 10.1080/19475705.2026.2637789.
- [11] L. Li, Y. Wang, J. Zhang, J. Tang, F. Yu, J. Wang. And Z. Zhang, "Characteristics of Ionospheric Responses over China During the November 2023 Geomagnetic Storm and Evaluation of Positioning Performance of CORS in Low-Latitude Regions," *Sensors (Basel)*, vol. 26, no. 7, p. 2198, pp. 1–23, Apr. 2026, doi: 10.3390/S26072198.
- [12] K. Wirsing and K. Wirsing, "Time Frequency Analysis of Wavelet and Fourier Transform in *Wavelet Theory*, IntechOpen Limited, London: United Kingdom. 2020, doi: 10.5772/INTECHOPEN.94521.
- [13] T. Du Nguyen and P. D. Nguyen, "Improvements in the Wavelet Transform and Its Variations: Concepts and Applications in Diagnosing Gearbox in Non-Stationary Conditions," *Applied Sciences*, vol. 14, pp. 1–16, May. 2024, doi: 10.3390/AP14114642.
- [14] L. Debnath, "Wavelet Transforms and Time-Frequency Signal Analysis," *Applications of Mathematics*, vol. 48, no. 1, pp. 79, Mar. 2003, doi: 10.1198/TECH.2002.S660.
- [15] J. R. K. Kumar Dabbakuti and D. Venkata Ratnam, "Characterization of Ionospheric Variability in TEC using EOF And Wavelets Over Low-Latitude GNSS Stations," *Advances in Space Research*, vol. 57, no. 12, pp. 2427–2443, Jun. 2016, doi: 10.1016/J.ASR.2016.03.029.
- [16] A. Krankowski, I. I. Shagimuratov, L. W. Baran, and I. I. Ephishov, "Study of TEC Fluctuations in Antarctic Ionosphere During Storm Using GPS Observations," *Acta Geophysica Polonica*, vol. 53, no. 2, pp. 205–218, Jan. 2005.
- [17] O. Mandrikova, Y. Polozov and B. Mandrikova, "Natural Data Analysis Method Based on Wavelet Filtering and NARX Neural Networks," *Engineering Proceedings*, vol. 33, pp. 1–8, Aug. 2023, doi: 10.3390/ENGPROC2023033063.
- [18] Y. Wang, L. Han, W. Liu, S. Yang and Y. Gao, "Study on Wavelet Neural Network Based Anomaly Detection in Ocean Observing Data Series," *Ocean Engineering*, vol. 186, pp. 1–9, Aug. 2019, doi: 10.1016/J.OCEANENG.2019.106129.
- [19] O. V. Mandrikova, N. V. Fetisova, Y. A. Polozov, I. S. Solovev, and M. S. Kupriyanov, "Method for Modeling of The Components of Ionospheric Parameter Time Variations and Detection of Anomalies in The Ionosphere," *Earth, Planets and Space*, vol. 67, no. 1, pp. 1–16, Aug. 2015, doi: 10.1186/S40623-015-0301-4.

- [20] H. Zhang and H. Zhang, "Anomaly Detection and Objective Security Evaluation Using Autoencoder, Isolation Forest, and Multi-Criteria Decision Methods," *Sensors (Basel)*, vol. 25, no. 19, pp. 1–16, Oct. 2025, doi: 10.3390/S25196250.
- [21] M. S. Binetti, V. F. Uricchio, and C. Massarelli, "Isolation Forest for Environmental Monitoring: A Data-Driven Approach to Land Management," *Environments*, vol. 12, no. 4, Page 1–18, Apr. 2025, doi: 10.3390/ENVIRONMENTS12040
- [22] F. Angiulli, F. Fassetti, and L. Ferragina, "Reconstruction Error-Based Anomaly Detection with Few Outlying Examples," *Neurocomputing*, vol. 675, pp. 1–20, Apr. 2026, doi: 10.1016/J.NEUCOM.2026.133002.
- [23] A. Luthfi and E. Emigawaty, "Isolation Forest-Based Anomaly Detection in IoT Smart Home Network Traffic," *ILKOM Jurnal Ilmiah*, vol. 18, no. 1, pp. 43–57, Apr. 2026, doi: 10.33096/ILKOM.V18I1.3156.43-57.
- [24] P. K. Mohanty, S. A. J. Francis, R. K. Barik, D. S. Roy, and M. J. Saikia, "Leveraging Shapley Additive Explanations for Feature Selection in Ensemble Models for Diabetes Prediction," *Bioengineering*, vol. 11, no. 12, pp. 1–19, Nov. 2024, doi: 10.3390/BIOENGINEERING11121215.
- [25] A. Gangwal and A. Lavecchia, "Explainable AI methods for drug discovery: A survey of interpretability, metrics and mechanistic insight," *Comput. Sci. Rev.*, vol. 61, pp. 1–25, Aug. 2026, doi: 10.1016/J.COSREV.2026.100943.
- [26] M. R. Santos, A. Guedes, and I. Sanchez-Gendríz, "SHapley Additive exPlanations (SHAP) for Efficient Feature Selection in Rolling Bearing Fault Diagnosis," *Machine Learning and Knowledge Extraction 2024*, vol. 6, no. 1, pp. 316–341, Feb. 2024, doi: 10.3390/MAKE6010016.
- [27] Y. Jin and K. Oksavik, "GPS Scintillations and Losses of Signal Lock at High Latitudes During the 2015 St. Patrick's Day Storm," *J. Geophys. Res. Space Phys.*, vol. 123, no. 9, pp. 7943–7957, Sep. 2018, doi: 10.1029/2018JA025933.
- [28] V. B. Belakhovsky, Y. Jin, and W. J. Miloch, "Influence of Different Types of Ionospheric Disturbances on GPS Signals At Polar Latitudes," *Ann. Geophys.*, vol. 39, no. 4, pp. 687–700, Jul. 2021, doi: 10.5194/ANGE0-39-687-2021.
- [29] "GNSS scintillation and TEC data at Skibotn in 2014–2023 - University of Oslo." Accessed: Jul. 28, 2026. [Online]. Available: <https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/RSBWOH>
- [30] R. V. Poursorkhabi, A. R. Fard, and A. R. Fard, "Designing A Model for Earthquake Timing and Magnitude Prediction Based on Neural Networks and Particle Swarm Optimization (PSO) algorithm," *Discover Geoscience 2026 4:1*, vol. 4, no. 42, pp. 1–29, Jan. 2026, doi: 10.1007/S44288-026-00427-3.
- [31] S. Goswami, S. Ghosh, and A. Paul, "Multifrequency Observations of Postmidnight Ionospheric Irregularities from an Anomaly Crest Location," *Radio Sci.*, vol. 57, no. 8, pp. 1–16, Aug. 2022, doi: 10.1029/2022RS007437;ISSUE:ISSUE:DOI.
- [32] S. Sunda, B. M. Vyas, and P. V. Khekale, "Storm Time Spatial Variations in TEC During Moderate Geomagnetic Storms in Extremely Low Solar Activity Conditions (2007–2009) over Indian region," *Advances in Space Research*, vol. 52, no. 1, pp. 158–176, July 2013, doi: 10.1016/J.ASR.2013.03.006.
- [33] K. Furmańczyk and M. Dudziński, "Outlier Detection Methods for One-Dimensional Features Based on the Isolation Forest and 2-Means Algorithms," *Model Assisted Statistics and Applications*, vol. 21, no. 3, pp. 1–17, June. 2026, doi: 10.1177/15741699261456599
- [34] H. Y. Gao, D. H. Zhang, Z. Z. Liu, S. J. Sun, Y. Q. Hao, and Z. Xiao, "Revisiting the Variation of the Ionospheric Irregularities in the Low Latitude Region of China Based on Small Regional Geodetic GNSS Station Network," *Space Weather*, vol. 21, no. 8, pp. 1–19, Aug. 2023, doi: 10.1029/2023SW003452;CTYPE:STRING:JOURNAL.
- [35] S. Karatay, A. Cinar, F. Arikan, and O. Arikan, "Spectral Tracks in the Ionosphere: Fast Fourier Transform Analysis Of Seismo-Ionospheric Coupling Through GPS Total Electron Content Variations," vol. 933, pp. 1–26, June. 2026, doi: 10.21203/RS.3.RS-7186871/V1.
- [36] R. H. Linjouom Nsangou and Y. Reuveni, "Global Ionospheric TEC Variability: Long-Term Spectral–Wavelet Analysis and Geomagnetic Storm Event Case Studies," *J. Geophys. Res. Space Phys.*, vol. 131, no. 4, pp. 1–17. e2025JA035005, Apr. 2026, doi: 10.1029/2025JA035005