Earth Science Informatics (2026) 19:62
https://doi.org/10.1007/s12145-026-02109-7

RESEARCH

AdaBoost-powered multi-class classification of pre-earthquake ionospheric anomalies using GNSS network in Türkiye: A comparison with random forest

Secil Karatay[1] · Zeynep Mantaroglu[1] · Esra Nur Serbetli[1]

Received: 30 October 2025 / Accepted: 16 March 2026 / Published online: 10 April 2026

Abstract

This study presents a novel Machine Learning framework for detecting pre-earthquake ionospheric anomalies using the Adaptive Boosting (AdaBoost) ensemble algorithm, applied to high-resolution Total Electron Content (TEC) data derived from Türkiye's dense TNPGN-Active GNSS network. Within the Lithosphere-Atmosphere-Ionosphere Coupling (LAIC) paradigm, we address key challenges in earthquake precursor research by implementing a three-class classification scheme to distinguish genuine seismo-ionospheric disturbances (Class-2: three days preceding earthquakes) from baseline variability under geomagnetically quiet (Class-0) and disturbed (Class-1) conditions, while incorporating geomagnetic indices (Kp, Ap, Dst) to filter space weather effects. IONOLAB-TEC estimates are analyzed for three M≥6.0 earthquakes in Türkiye: Elazığ (Mw 6.7, 24 January 2020), İzmir (Mw 7.0, 30 October 2020) and the Kahramanmaraş doublet (Mw 7.8 and 7.5, 6 February 2023). For each event, TEC vectors from 12 strategically selected stations are Z-score normalized and classified using AdaBoost with decision trees as weak learners (T = 100 iterations, $learning_{rate}$=1.0, MaxNumSplits=20), evaluated under 10-fold cross-validation with training set sizes varied from 10% to 100% in 5% increments. The model achieves Cohen's kappa values ranging from 0.744 to 0.992 (substantial to almost perfect agreement), with classification Accuracy varying from 82.94% to 99.48% depending on earthquake characteristics. Performance varies regionally: Elazığ demonstrates exceptional results (99.48% Accuracy, κ = 0.992), Kahramanmaraş shows strong performance (90.21% Accuracy, κ = 0.853), while İzmir exhibits more modest results (82.94% Accuracy, κ = 0.744). Pre-earthquake class (Class-2) metrics vary across tectonic settings, with detailed per-class results provided in Supplementary Table S1. Feature importance analysis reveals distributed spatial sampling with algorithm-dependent inverse correlation between station importance and epicentral distance. AdaBoost shows weak-to-moderate correlation (Spearman ρ = −0.55 to 0.10, generally non-significant), while Random Forest captures this spatial relationship more consistently (ρ = −0.37 to −0.69, with E3-E4 reaching significance at p = 0.04). Statistical validation (McNemar's test, p < 0.001; Cohen's kappa up to 0.992 for AdaBoost and 1.0 for Random Forest) outperforms binary baselines, highlighting the robustness of both ensemble methods in handling noisy, imbalanced geophysical data. Rigorous 10-fold cross-validation confirms that Random Forest consistently outperforms AdaBoost across all tectonic settings (E1: 100.00% vs. 99.49%; E2: 99.78% vs. 83.47%; E3-E4: 99.69% vs. 88.33%, p < 0.001), establishing Random Forest as the preferred algorithm for operational seismo-ionospheric monitoring. Both ensemble methods substantially exceed traditional threshold-based approaches, validating Machine Learning's potential for earthquake precursor detection. This study demonstrates the critical importance of cross-validation with fold-wise normalization for small geophysical datasets, as proper validation protocols are essential to ensure reported performance reflects genuine generalization rather than methodological artifacts.

Keywords Seismo-ionospheric precursors · Remote sensing · Total Electron Content (TEC) · Machine Learning · AdaBoost ensemble · GNSS · Geomagnetic activity filtering

Communicated by Hassan Babaie

Extended author information available on the last page of the article

Introduction

The investigation of seismo-ionospheric coupling mechanisms has emerged as a central focus in contemporary earthquake precursor research. This interest is driven by its profound implications for enhancing seismic forecasting capabilities. The Lithosphere-Atmosphere-Ionosphere Coupling (LAIC) framework has gained widespread acceptance as a comprehensive theoretical model. It proposes that stress buildup in the lithosphere preceding $M \geq 6$ earthquakes can manifest as detectable anomalies in both atmospheric and ionospheric parameters (Pulinets et al. 1994; Pulinets and Herrera 2024; Pulinets and Ouzounov 2011; Karatay et al. 2010; Pirti and Karatay 2025). The LAIC mechanism functions through three fundamental pathways: (i) a thermal/chemical channel involving enhanced Radon and volatile gas emissions that modify air ionization patterns and surface thermal radiation (influencing latent heat flux and air conductivity) (Woith 2015; Yaşar et al. 2025); (ii) an Acoustic Gravity Wave (AGW) channel through which crustal fracturing generates upward-propagating AGWs (driving mechanical energy transfer and atmospheric density oscillations) (Srivastava et al. 2021) and (iii) an electromagnetic channel whereby seismogenic electric fields penetrate atmospheric layers and reach the ionosphere (inducing ionospheric plasma drift and electron density redistribution) (De Santis et al. 2019; Pulinets et al. 2022). Over recent decades, accumulating evidence supports the hypothesis that M≥6.0 earthquakes may be preceded by anomalous ionospheric variations, presenting a promising avenue for earthquake prediction research (Karatay et al. 2010; Astafyeva 2019; Shah et al. 2020; Karatay et al. 2025). Multi-platform observational data from satellite-based and ground-based measurements of thermal radiation, atmospheric parameters and electromagnetic disturbances have substantially contributed to the growing body of evidence supporting LAIC-related phenomena. Despite ongoing scientific discourse regarding the precise physical mechanisms, LAIC anomalies are increasingly recognized as reliable indicators of seismic activity. Within the LAIC theoretical framework, pre-seismic ionospheric disturbances manifest through variations in Total Electron Content (TEC). These disturbances have been documented in temporal windows ranging from days to hours before major events and are observed in association with numerous global $M \geq 6$ earthquakes (Budak et al. 2024; Heki et al. 2024; Ullah et al. 2025).

The ionosphere a dynamically evolving plasma layer extending approximately 60–1000 km above Earth's surface, exhibits responsiveness to both terrestrial and extra-terrestrial forcing mechanisms. While solar activity and geomagnetic variations constitute the primary drivers of ionospheric variability, mounting evidence indicates that lithospheric processes preceding M≥6.0 earthquakes can induce observable ionospheric variations (Shen et al. 2018; Marrissey et al. 2020). Within the electromagnetic channel of LAIC, seismogenic electric fields manifest through multiple ionospheric signatures with TEC variations representing one of the most reliably measured parameters using Global Navigation Satellite System (GNSS) observations (Xie et al. 2021; Ullah et al. 2025; Karatay et al. 2025). Although comprehensive understanding remains incomplete, these seismo-ionospheric coupling mechanisms likely involve multiple interconnected pathways influenced by pre-seismic rock deformation and stress accumulation. Conventional methodologies for detecting ionospheric disturbances have predominantly relied on statistical analyses and threshold-based approaches. However, these methods frequently encounter difficulties in distinguishing seismic-induced anomalies from natural ionospheric variability (Liu et al. 2001, 2004a, b, 2018; Joshi et al. 2023; Zhang et al. 2025). The ionosphere's complex temporal and spatial patterns, shaped by diurnal cycles, seasonal variations, solar activity cycles and geomagnetic conditions, present substantial challenges for accurate earthquake precursor identification. Furthermore, the nonlinear and potentially chaotic nature of seismo-ionospheric interactions necessitates analytical approaches capable of capturing subtle patterns that conventional statistical methods may overlook. The need to differentiate between ionospheric disturbances caused by geomagnetic storms and those potentially related to seismic activity remains a critical challenge in this research domain.

Machine Learning (ML) methodologies represent a paradigm shift in addressing these analytical challenges. Supervised learning methods are particularly valuable for this application because they can capture complex nonlinear relationships and feature interactions that traditional threshold-based approaches cannot model, while providing probabilistic confidence estimates for classification decisions (Karatay and Gul 2023; Ridzwan and Yusoff 2023; Convertito et al. 2024). Recent advances in ML algorithms, enhanced computational capabilities and the increasing availability of multi-source geophysical datasets have created unprecedented opportunities for detecting, characterizing and validating earthquake-related ionospheric anomalies. Throughout the past decade, ML applications in seismo-ionospheric studies have progressed rapidly. Initial implementations primarily employed supervised learning techniques, including Support Vector Machines (SVM) (Akyol et al. 2020; Asaly et al. 2022), K-Nearest Neighbors (Budak et al. 2024), Neural Networks (Lin 2022; Uyanık et al. 2023), Random Forest (Zorkun et al. 2025) for classifying predefined ionospheric features. More sophisticated approaches have recently incorporated Deep Learning (DL) architectures, including Convolutional Neural Networks (CNNs) (Shan etal. 2022; Uyanık et al. 2023), Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks (Berhich et al. 2022; Kiruthiga and Mythili 2024; Mukeshet al. 2024), which have demonstrated particular promise in capturing spatial, temporal and multi-scale characteristics of ionospheric perturbations (Xiong et al. 2022; Zhang et al. 2025). A significant advancement in this research domain is the integration of ML techniques with multi-source observational data. Contemporary studies increasingly combine ionospheric observations from GNSS networks with complementary geophysical parameters, including geomagnetic indices (Kp, Ap, Dst), to better distinguish seismo-ionospheric anomalies from space weather effects (Karatay and Gul 2023; Li et al. 2024; Budak et al. 2024; Zorkun et al. 2025). The incorporation of geomagnetic activity indicators enables more reliable identification of earthquake-related ionospheric disturbances by filtering out variations attributable to solar-terrestrial interactions. The development of classification frameworks capable of distinguishing between quiet ionospheric conditions, geomagnetically disturbed periods and genuine pre-earthquake anomalies represents a critical step toward reliable earthquake precursor identification.

Among ensemble learning methods, boosting algorithms have demonstrated particular effectiveness in handling imbalanced datasets and improving classification Accuracy through iterative refinement. Adaptive Boosting (AdaBoost), originally developed by Freund and Schapire (1997), represents a powerful ensemble technique that combines multiple weak learners to create a strong classifier (Freund and Schapire 1997). The algorithm iteratively adjusts the weights of training instances, focusing subsequent learners on previously misclassified examples, thereby progressively improving overall performance. AdaBoost has been successfully applied across various domains and in earthquake prediction studies, demonstrating robust performance in complex classification tasks (Asim et al. 2018; Kaushal et al. 2025). In geophysical applications, ensemble methods have shown promise for detecting subtle anomalies in multi-dimensional datasets, making them particularly suitable for seismo-ionospheric precursor identification. AdaBoost offers computational efficiency, interpretability through feature importance analysis and robust performance with limited training data (108 samples per earthquake), with outlier sensitivity mitigated through its weighting mechanism (Drucker 1997; Saied et al. 2023). While AdaBoost demonstrates strong performance on many geophysical datasets, its effectiveness for seismo-ionospheric classification with limited samples (N=108 per earthquake) requires systematic validation through rigorous cross-validation and comparison with alternative ensemble methods such as Random Forest (RF). Direct comparison with RF on identical raw TEC vectors (Sect. 5.4, Tables 5 and 6) demonstrates the critical importance of cross-validation for small geophysical datasets, as single train-test splits produced misleading conclusions that cross-validation corrected.

Türkiye is located in one of the most seismically active regions globally, where the complex interaction of multiple tectonic plates, including the North Anatolian Fault (NAF) and East Anatolian Fault (EAF) systems, generates frequent and destructive earthquakes (Aktug et al. 2016). The country's geographical position at the convergence of the Arabian, African and Eurasian plates creates intense tectonic stress accumulation, resulting in devastating seismic events that have caused significant loss of life and economic damage throughout history. In recent years, Türkiye has experienced several M≥6.0 earthquakes that have highlighted the urgent need for improved earthquake forecasting capabilities. The Elazığ earthquake (Mw 6.7) on January 24, 2020, struck eastern Türkiye along the East Anatolian Fault. Later the same year, on October 30, 2020, the İzmir earthquake (Mw 7.0) occurred in the Aegean Sea region, severely affecting western Türkiye and Greece. Most recently, on February 6, 2023, two catastrophic earthquakes (Mw 7.8 and 7.5) struck Kahramanmaraş and surrounding provinces, marking one of the deadliest natural disasters in Türkiye's modern history with more than 50,000 fatalities and widespread destruction (Barbot et al. 2023; Melgar et al. 2023; Solís-Alulima et al. 2023). These events underscore the critical importance of advancing precursor-based early warning systems in Türkiye, particularly through the integration of multi-parameter geophysical observations and ML techniques. The dense network of seismic, Global Positioning System (GPS) and ionospheric monitoring stations across the country provides a unique opportunity to detect subtle seismo-ionospheric and strain anomalies prior to major ruptures (Oyama et al. 2016; Haider et al. 2025). Leveraging such datasets with robust ensemble classifiers like AdaBoost can enhance the reliability of short-term earthquake forecasting, ultimately contributing to disaster risk reduction in this high-hazard region.

This paper introduces a framework for detecting earthquake-related ionospheric disturbances using the AdaBoost ensemble learning algorithm. The proposed methodology addresses critical challenges in the field, including distinguishing seismic precursors from solar-geomagnetic influences through multi-class classification and quantifying detection uncertainty through comprehensive performance metrics. To evaluate the effectiveness of AdaBoost for this application, we compare its performance with RF, another powerful ensemble method, applied to the same dataset by our research group (Zorkun et al. 2025), demonstrating the complementary strengths of different ML architectures for seismo-ionospheric precursor detection. Both single train-test split and 10-fold cross-validation consistently demonstrate RF's superiority across all tectonic settings. Single-split results (Table 5) show RF achieving 100.00%, 99.87% and 99.54% Accuracy for E1, E2 and E3-E4 respectively, compared to AdaBoost's 99.48%, 82.94% and 90.21%. Cross-validation (Table 6) confirms these findings with statistical rigor: RF achieves 100.00 ± 0.00%, 99.78 ± 0.38% and 99.69 ± 0.33% versus AdaBoost's 99.49 ± 0.24%, 83.47 ± 2.09% and 88.33 ± 1.53% ($p < 0.001$ for E2 and E3-E4). This consistency across evaluation strategies validates the robustness of the comparison, while demonstrating that cross-validation with fold-wise normalization is essential practice for seismo-ionospheric ML studies with limited samples ($N = 108$).

Despite AdaBoost's computational efficiency advantage (0.69–0.74 s vs. 1.57–1.79 s training time), RF emerges as the preferred algorithm for operational monitoring due to superior generalization and substantially lower performance variance. The framework's effectiveness is evaluated using a comprehensive dataset encompassing three M ≥ 6.0 earthquakes that occurred in Türkiye between 2020 and 2023: the Elazığ earthquake, the İzmir earthquake and the Kahramanmaraş earthquakes. The analysis employs a three-class classification scheme to differentiate between: (1) three geomagnetically quiet days (Class-0), representing baseline ionospheric conditions; (2) three geomagnetically disturbed days (Class-1), corresponding to periods of enhanced space weather activity and (3) three pre-earthquake days (Class-2), encompassing the temporal windows immediately preceding each seismic event. This classification framework enables systematic evaluation of the model's ability to distinguish genuine earthquake precursors from background ionospheric variability and space weather effects. GNSS-derived TEC estimates from Turkish National Permanent GNSS Network-Active (TNPGN-Active) stations constitute the primary input data for this study. The network provides dense spatial coverage across Türkiye's seismically active regions, enabling detailed characterization of ionospheric variations in the vicinity of earthquake epicenters. For each earthquake, TEC data are collected from multiple GNSS stations strategically positioned to capture potential pre-seismic ionospheric anomalies while providing sufficient spatial diversity for robust statistical analysis. The study incorporates geomagnetic activity indices, Kp, Ap and Dst to facilitate differentiation between ionospheric disturbances induced by space weather events and those potentially associated with seismic processes.

The AdaBoost classifier uses 70/30 train-test split with Z-score normalization, evaluated through Accuracy, Precision, Recall and F1-score metrics. The methodology and findings are presented in subsequent sections: Sect. 2 describes data sources, Sect. 3 details the classification methodology and Sects. 4–5 present results and interpretation.

Description of the data, study area and space weather conditions

In this study, GPS-derived Total Electron Content (TEC) is computed using the Regularized Estimation of TEC (Reg-Est) method, which leverages dual-frequency measurements from Continuously Operating Reference Stations (CORS), as detailed in Arikan et al. (2003, 2004) and Nayir et al. (2007). Reg-Est is an advanced signal processing technique that mitigates the ill-posed nature of TEC inversion by applying spatial and temporal regularization, yielding more accurate and stable TEC estimates from dual-frequency GNSS observations compared to traditional methods. This approach is selected for its proven robustness and effectiveness in managing noisy or incomplete data. The IONOLAB-TEC values utilized here have a temporal resolution of 2.5 min, generating 576 estimates per day for each station. This high temporal resolution is particularly advantageous for capturing subtle, short-term seismo-ionospheric variations and high-frequency precursor signals that might be averaged out or lost in lower-resolution datasets. The unit of TEC is TEC Units (TECU), where $1\,TECU = 10^{16}$ electrons/m² (Arikan et al. 2003, 2004; Nayir et al. 2007; Sezen et al. 2013; IONOLAB, 2006).

The TNPGN-Active stations included in this analysis are chosen based on the following criteria: (1) geographic coverage within approximately 1000 km of earthquake epicenters to capture spatial variations in ionospheric perturbations, (2) stations positioned within or near the estimated Earthquake Preparation Zones (EPZs, calculated using Dobrovolsky et al. 1979 formula: $R = 10^{0.43M}$), (3) data completeness exceeding 95% throughout the study periods with < 5% missing observations and (4) active operational status within the TNPGN-Active network during 2020–2023. We acknowledge that formalized, quantitative station selection criteria (e.g., optimal density thresholds, distance-weighting schemes) remain a methodological gap that future work should address through systematic sensitivity analyses.

Recent studies have demonstrated the effectiveness of GNSS-derived TEC for detecting co-seismic and pre-seismic ionospheric disturbances (Maletckii et al. 2023; Vesnin et al. 2023; Heki et al. 2024). Maletckii et al. (2023) and Vesnin et al. (2023) reported clear N-wave signatures in TEC data following the 2023 Türkiye earthquake sequence with co-seismic ionospheric disturbances propagating from the epicentral areas at velocities of 750–1100 m/s. Heki et al. (2024) identified pre-seismic TEC anomalies occurring within hours to days before M≥6.0 earthquakes, consistent with the LAIC mechanism. These findings support the utilization of high-resolution TEC time series for investigating seismo-ionospheric phenomena in the context of earthquake precursor research.

The three analyzed earthquakes (Table 1) represent diverse Turkish seismotectonic settings: continental strike-slip on NAF/EAF (Elazığ, Kahramanmaraş) and offshore normal faulting in the Aegean extensional regime (İzmir). The Elazığ earthquake (E1-Mw 6.7) struck eastern Türkiye along the East Anatolian Fault on January 24, 2020, causing severe destruction in Elazığ and Malatya provinces. Later that year, on October 30, 2020, the İzmir earthquake (E2-Mw 7.0) occurred in the Aegean Sea, impacting western Türkiye and Greece, with particularly heavy damage in İzmir's Bayraklı district. Most recently, on February 6, 2023, two powerful earthquakes (E3 and E4-Mw 7.8 and Mw 7.5) ravaged Kahramanmaraş and adjacent provinces, marking one of the deadliest disasters in Türkiye's modern history. Detailed information on these earthquakes is provided in Table 1 (USGS, 1879). The epicenters of three earthquakes and the chosen TNPGN-Active stations are presented with red stars and blue points in Fig. 1, respectively. The geographic coordinates of these stations and earthquakes in which the stations were used are given in Table 2.

In this study, IONOLAB-TEC data from 12 TNPGN-Active stations are utilized for each of the three earthquake events analyzed. The dataset is constructed to distinguish pre-earthquake ionospheric anomalies from normal ionospheric variations under different geomagnetic conditions. The IONOLAB-TEC estimates are categorized into three hierarchical classes based on the geomagnetic state and temporal relationship to earthquake occurrence:

Class-0 (Geomagnetically quiet background): IONOLAB-TEC values from three geomagnetically quiet days are selected for each earthquake event. These days are characterized by minimal geomagnetic disturbances and represent the baseline ionospheric behavior under quiet space weather conditions. Class-0 data serve as the reference background for normal ionospheric variability.

Class-1 (Geomagnetically disturbed background): Three geomagnetically disturbed days are identified for each event, during which the ionosphere exhibited enhanced variability due to increased geomagnetic activity. These data capture IONOLAB-TEC variations induced by space weather phenomena and represent ionospheric perturbations unrelated to seismic activity.

Class-2 (Pre-earthquake anomalies): IONOLAB-TEC estimates from the three days immediately preceding each earthquake are designated as Class-2. These data potentially contain seismo-ionospheric precursors and constitute the target class for anomaly detection.

For the Kahramanmaraş doublet (E3 and E4, separated by 9 h), the events are treated as a single earthquake sequence with a unified pre-earthquake period (3–5 February 2023). This combined treatment reflects the overlapping preparation windows, the first rupture (E3, 01:17 UTC) occurred within the preparation phase of the second (E4, 10:24 UTC),

Table 1 Date, occurring time, geographic locations, magnitude, depth and EPZ (Dobrovolsky et al. 1979) of the chosen earthquakes. All earthquakes occurred within Türkiye

Earthquake	Date (dd.mm.yyyy)	Time (UT)	Latitude (°*N*)	Longitude (°E)	Magnitude (Mw)	Depth (km)	EPZ (km)
E1	24.01.2020	17.55	38.37	39.10	6.7	10	760
E2	30.10.2020	11.51	37.89	26.79	7.0	21	1023
E3	06.02.2023	01.17	37.11	37.11	7.8	17.9	2259
E4	06.02.2023	10.24	38.02	37.20	7.5	10	1679

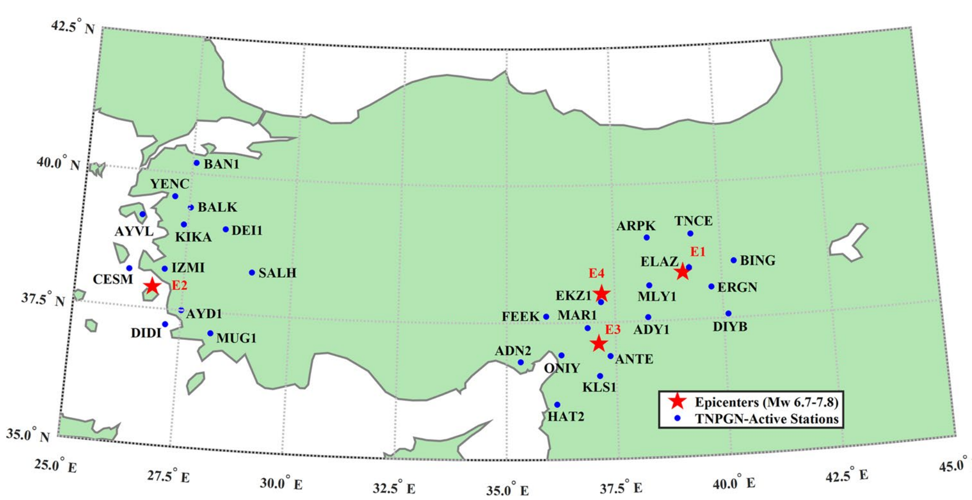


Fig. 1 The geographic coordinates of three epicenters and the chosen TNPGN-Active stations used in the study

Table 2 The geographic locations of the stations and earthquakes in which the stations were used. All earthquakes occurred within Türkiye

Station name (station ID)	Latitude (°*N*)	Longitude (°E)	Earthquake
Adana (ADN2)	36.79	35.32	E1, E3 and E4
Adıyaman (ADY1)	37.57	38.26	E1, E3 and E4
Arapgir (ARPK)	39.02	38.29	E1
Aydın (AYD1)	37.50	27.50	E2
Ayvalık (AYVL)	39.18	26.41	E2
Balıkesir (BALK)	39.38	27.53	E2
Bandırma (BAN1)	40.20	27.58	E2
Bingöl (BING)	38.53	40.30	E1
Çeşme (CESM)	38.18	26.22	E2
Denizli (DEI1)	39.03	28.39	E2
Didim (DIDI)	37.22	27.16	E2
Diyarbakır (DIYB)	37.57	40.11	E1
Ekinözü (EKZ1)	37.87	37.18	E1, E3 and E4
Elazığ (ELAZ)	38.45	39.25	E1, E3 and E4
Ergani (ERGN)	38.08	39.75	E3 and E4
Feke (FEEK)	37.62	35.91	E3 and E4
Gaziantep (ANTE)	36.88	37.37	E3 and E4
Hatay (HAT2)	36.01	36.14	E1, E3 and E4
İzmir (IZMI)	38.23	27.04	E2
Kahramanmaraş (MAR1)	37.40	36.86	E1, E3 and E4
Kırkağaç (KIKA)	39.06	27.40	E2
Kilis (KLS1)	36.52	37.12	E3 and E4
Malatya (MLY1)	38.15	38.31	E1, E3 and E4
Muğla (MUG1)	37.12	28.21	E2
Osmaniye (ONIY)	36.91	36.25	E1, E3 and E4
Salihli (SALH)	38.28	29.07	E2
Tunceli (TNCE)	39.06	39.32	E1
Yenice (YENC)	39.56	27.14	E2

Table 3 The chosen periods for each classes representing the space weather conditions

	Pre-earthquake period	Disturbed period	Quiet period
		Class	
	2	1	0
E1	21–23.01.01.2020	23–25.04.04.2023	17–19.02.02.2022
E2	27–29.10.10.2020	23–25.04.04.2023	17–19.02.02.2022
E3	03–05.02.02.2023	23–25.04.04.2023	17–19.02.02.2022

meaning that ionospheric anomalies during this period likely reflect cumulative stress buildup from both impending events rather than isolated precursors.

The selection of geomagnetically quiet (Class-0) and disturbed (Class-1) days is performed based on standard space weather indices, including the Kp (planetary K-index), Ap (planetary amplitude index) and Dst (Disturbance storm time index). These indices enabled the systematic identification of periods with varying levels of geomagnetic activity. The identified periods for each class are given in Table 3. The variations in Dst, Kp and Ap indices in the periods determined for each class are also shown in Fig. 2.

Following data classification, the AdaBoost ensemble learning algorithm is applied to perform multi-class classification, enabling the discrimination of pre-earthquake ionospheric signatures (Class-2) from background variations under both quiet (Class-0) and disturbed (Class-1) geomagnetic conditions.

Geomagnetic storm days are selected based on data from the Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences (IZMIRAN, 1939). Earthquake data is sourced from the United States Geological Survey at https://earthquake.usgs.gov/earthquakes (USGS, 1879). The Kp and Ap indices are retrieved from the GFZ Helmholtz Centre for Geosciences (GFZ, 1956), while the Dst index is obtained from the World Data Center for Geomagnetism at (WDC, 1999).

Analysis of space weather conditions during the study periods revealed that the Elazığ earthquake (January 2020) occurred during relatively quiet geomagnetic conditions with average Kp values below 3 and Dst fluctuations within ±30 nT. The İzmir earthquake (October 2020) period similarly exhibited predominantly quiet conditions, though moderate enhancements in Kp (reaching 4–5) are observed 5–7 days before the event due to a minor geomagnetic storm. The Kahramanmaraş earthquake sequence (February 2023) occurred during an extended period of low solar activity with geomagnetic indices remaining at quiet levels (Kp≤2 and Dst > −20 nT) throughout the two weeks preceding the earthquakes. These relatively quiescent space weather conditions are favorable for detecting earthquake-related ionospheric signatures, as they minimize contamination from geomagnetic storm-induced TEC variations. Note that E1 and E3/E4 occurred during relatively quiet geomagnetic conditions similar to the quiet baseline, whereas the E2 period exhibits moderate geomagnetic activity with elevated Kp and Ap values, complicating the isolation of seismo-ionospheric anomalies.

In this study, the severe geomagnetic storm of April 23–25, 2023 (reaching Kp values of 8), is deliberately selected as a common 'high-noise' benchmark for all analyzed earthquakes. While utilizing diverse storm events for each earthquake would enhance generalizability, employing a single, consistent, high-intensity disturbance period allows for a controlled comparison of the classifier's performance across different seismic events without the confounding variable of varying storm intensities. By testing all earthquakes against this same 'worst-case' space weather scenario, we isolate the model's sensitivity to seismic precursors relative to a fixed noise floor. The use of a single geomagnetic storm event (23–24 April 2023, Dst = −233 nT) as the common disturbed-day benchmark across all three earthquake sequences is a

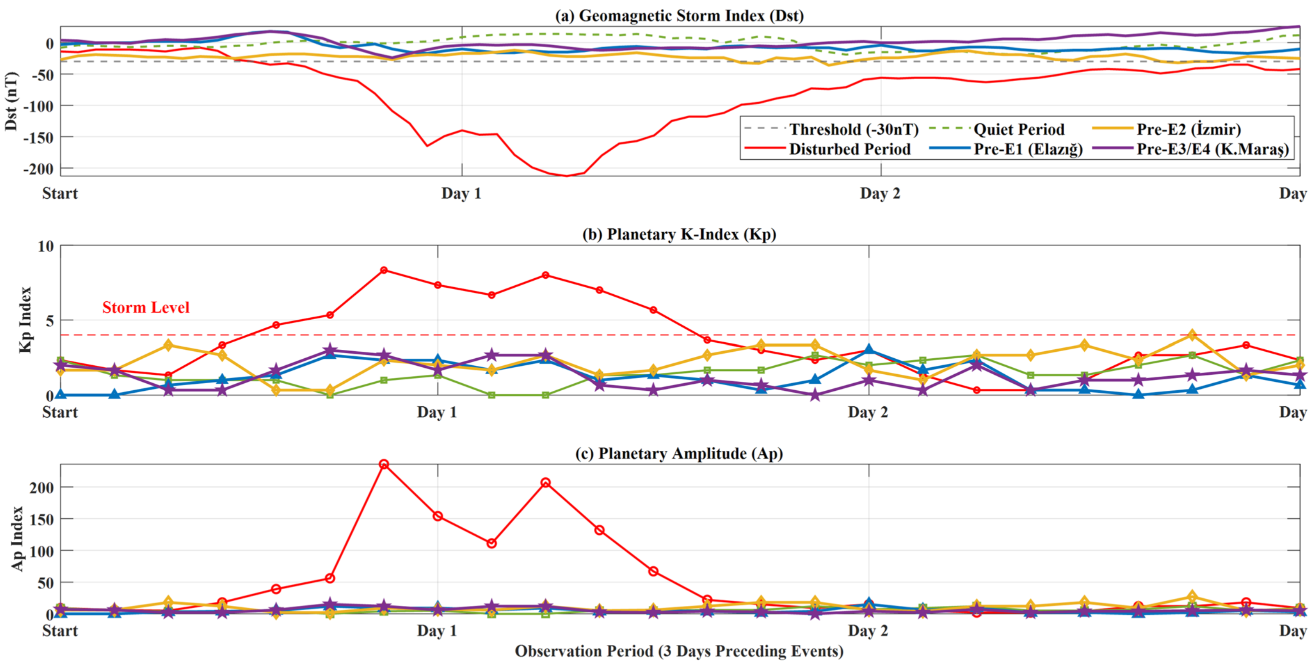


Fig. 2 Superposed epoch analysis of space weather conditions during the three-day observation windows. The panels illustrate the temporal variations of (**a**) hourly Dst (Disturbance storm time) index, (**b**) 3-hour planetary K-index (Kp) and (**c**) 3-hour planetary amplitude (Ap) index. The comparison includes the geomagnetically quiet baseline (17–19 Feb 2022), the disturbed baseline (23–25 Apr 2023) and the pre-seismic periods for Elazığ (E1), İzmir (E2) and Kahramanmaraş (E3/E4) earthquakes. The horizontal dashed lines indicate geomagnetic storm thresholds (Dst ≤ -30 nT and Kp$\geq$4)

deliberate design choice that isolates algorithm sensitivity under a fixed geomagnetic noise floor. This controlled approach ensures that performance differences across events reflect variations in pre-earthquake ionospheric signatures and tectonic settings rather than differences in disturbed-class characteristics. However, this design does not capture the full range of geomagnetic variability encountered in operational settings, including moderate storms, recurrent high-speed stream events and substorm-driven disturbances with distinct ionospheric morphologies. Future studies should incorporate multiple storm events spanning different intensities (-50 to -300 nT) and seasons to assess classifier robustness under varying geomagnetic noise conditions.

Implementation of adaptive boosting (AdaBoost) algorithm

While more advanced boosting variants like XGBoost or Gradient Boosting exist, AdaBoost is specifically selected for this study due to its computational efficiency in resource-constrained environments and its proven robustness when dealing with relatively small geophysical datasets (108 samples per event), where more complex models might be prone to overfitting. Furthermore, AdaBoost offers high interpretability through feature importance analysis, which is crucial for identifying the spatial contribution of GNSS stations to precursor detection.

The input data consist of IONOLAB-TEC vectors $\boldsymbol{x}_{u,d}$ from $U = 12$ GPS stations over D observation days. Each vector $\boldsymbol{x}_{u,d} = [x_{u,d}(1),\ x_{u,d}(2), \ldots, x_{u,d}(n), \ldots, x_{u,d}(N)]^T$ contains $N = 576$ TEC measurements per day (2.5-minute temporal resolution × 24 h) for station u on day d, forming a 576-dimensional feature vector. Each sample represents a single day's observations from one station. For each earthquake, the complete dataset comprises 108 samples: 12 stations × 3 days × 3 classes (quiet, disturbed, pre-earthquake), yielding 36 samples per class. Geomagnetic indices (Kp, Ap, Dst) are used exclusively for selecting Class-0 (quiet) and Class-1 (disturbed) reference days during data preparation (Sect. 2) and are not included as classifier input features, the AdaBoost model receives only Z-score normalized TEC vectors.

Each IONOLAB-TEC vector $\boldsymbol{x}_{u,d}$ is associated with a corresponding class label $c_{u,d} \in \{0,1,2\}$, defined as:

$$c_{u,d} = \begin{cases} 0, & \quad \quad \quad \quad \quad \quad \textit{Quiet day} \\ 1, & \quad \quad \quad \quad \quad \textit{Disturbed day} \\ 2, & \quad \textit{Pre}-\textit{earthquake day}. \end{cases} \tag{1}$$

The complete dataset D comprises 108 samples per earthquake ($12\ \textit{stations} \times\ 3\ \textit{days} \times\ 3\ \textit{classes}$).

All IONOLAB-TEC vectors are normalized using Z-score normalization to eliminate scale bias across features (Jain et al. 2000):

$$\widetilde{x}_{u,d}\left(n\right)=\frac{x_{u,d}\left(n\right)-\mu_{\ n}}{\sigma_{\ n}} \tag{2}$$

To prevent data leakage, normalization parameters are computed exclusively from the training data within each evaluation scheme: for the single 70/30 split, $\mu_{\ n}$ and $\sigma_{\ n}$ are derived solely from the training partition and applied to both training and test sets; for 10-fold cross-validation, $\mu_{\ n}$ and $\sigma_{\ n}$ are recomputed from the k−1 training folds at each iteration and applied to the held-out test fold. This fold-wise normalization ensures that no information from test samples influences the feature scaling, maintaining strict separation between training and evaluation data. The 108 samples are randomly shuffled and partitioned for model evaluation using 10-fold stratified cross-validation. To thoroughly evaluate model scalability and learning curves, the training sets within each fold are systematically subsampled from 10% to 100% in 5% increments. Each day constitutes an independent sample in both partitioning schemes.

For the single split, we use $random_{state}=42$ to ensure reproducibility and enable direct comparison between AdaBoost and RF on identical data partitions (addressing methodological concerns about confounding variables in algorithm comparison). However, given the limited sample size (N=108 per earthquake), reliance on a single train-test split is statistically unreliable and may produce misleading conclusions due to partition-specific bias. Therefore, we employ 10-fold stratified cross-validation as the primary performance evaluation method. In k-fold cross-validation, the dataset is divided into $k\ =\ 10$ equal-sized folds, with each fold serving once as the test set while the remaining k-1 folds form the training set. Stratification ensures proportional class representation (36 samples per class, or 12 samples per class per fold) across all folds, maintaining the original 1:1:1 class distribution. This process yields 10 independent Accuracy measurements per model, from which we compute mean and standard deviation to assess both average performance and generalization stability.

We note that the 108 station-day samples per earthquake exhibit a clustered structure: observations from different stations on the same calendar day share temporal ionospheric conditions, while observations from the same station across days share spatial characteristics. Standard k-fold cross-validation treats each station-day pair as independent, which may not fully account for this temporal and spatial correlation. Alternative validation strategies such as leave-one-day-out (partitioning by calendar day) or leave-one-station-out (partitioning by station) would provide more conservative generalization estimates by preventing correlated samples from appearing in both training and test sets. However, with only 9 unique days and 12 stations per earthquake, these approaches produce extremely small test folds (12 or 9 samples respectively) with high per-fold variance, limiting their statistical reliability. We therefore retain 10-fold stratified CV as the primary evaluation method while acknowledging that reported performance may be optimistic relative to fully independent temporal or spatial validation. This limitation should be addressed in future studies with expanded multi-event datasets enabling grouped cross-validation designs.

Statistical significance of performance differences between algorithms is assessed using paired t-tests across the 10 folds (Kohavi 1995; Dietterich 1998; Domingos 2012). All reported performance metrics in Sect. 5 (Accuracy, Precision, Recall, F1-score, Cohen's κ) represent cross-validated means unless explicitly noted as single-split results (Table 5). This dual reporting strategy allows readers to compare single-split versus cross-validated findings and understand the critical importance of rigorous validation for small geophysical datasets. To classify IONOLAB-TEC patterns into three geophysical states, the AdaBoost algorithm is employed, using decision trees as weak learners. (Freund and Schapire 1997; Breiman et al. 2017). Each training instance $\left(x_i,c_i\right)$ is assigned an equal initial weight:

$$w_i^{(1)}=\frac{1}{M_{train}} \tag{3}$$

where M_{train} is the number of training samples. The procedure is carried out iteratively for each time step $t=1,\ 2,\ \ldots,\ T$.

In each iteration, a decision tree $h_{\mathbf{t}}\left(\mathbf{x}\right)$ is trained on the weighted samples. The performance of this tree is measured by its weighted error. Weighted error can be defined as follows:

$$\epsilon_{\ t}=\frac{\sum_{m=1}^{M_{train}}w_m^{(t)}\mathbb{I}\left[h_t\left(x_m\right)\neq\ c_m\right]}{\sum_{\ m=1}^{\ M_{train}}w_m^{(t)}} \tag{4}$$

where $\mathbb{I}[\bullet]$ is the indicator function. This equation computes the weighted classification error, representing the sum of weights of the instances that the current weak learner misclassified. It quantifies how poorly the individual decision tree performed on the current distribution of data. The learner weight (importance factor) can be defined with Eq. (5):

$$\alpha_{\ t}=\eta\ \cdot\ \ln\left(\frac{1-\epsilon_{\ t}}{\epsilon_{\ t}}\right) \tag{5}$$

where $\eta = 1.0$ the default $learning_{rate}$ is applied. This formula calculates the learner weight (importance factor), which determines the influence of the current weak learner in the final ensemble. A lower ϵ_t leads to a higher α_t, giving more "voting power" to more accurate trees. In the computational implementation, the $learning_{rate}$ is applied as a shrinkage factor that scales α_t during weight updates, providing regularization to prevent overfitting (Freund and Schapire 1997; Hastie et al. 2009). Sample weights are updated as follows:

$$w_i^{(t+1)} = w_i^{(t)} \cdot \exp\left(\alpha_t \mathbb{I}\left[h_t\left(x_i\right) \neq c_i\right]\right) \tag{6}$$

followed by normalization:

$$w_i^{(t+1)} \leftarrow \frac{w_i^{(t+1)}}{\sum_{i=1}^{M_{train}} w_i^{(t+1)}} \tag{7}$$

Equations (6) and (7) perform the sample weight update and normalization. The algorithm increases the weights of misclassified samples, forcing the next weak learner to focus on these difficult geophysical patterns.

The AdaBoostM2 algorithm is implemented using an ensemble learning framework with decision trees as base learners. The iterative weight update process described in Eqs. (3)-(7) is handled by the ensemble training procedure, which automatically manages sample reweighting, weak learner training and ensemble aggregation across all boosting iterations. After $T = 100$ iterations, the final prediction function is expressed as follows:

$$H\left(X\right) = \underset{k \in \{0,1,2\}}{argmax} \sum_{t=1}^{T} \alpha_t \mathbb{I}\left[h_t\left(x\right) = k\right] \tag{8}$$

Equation (8) represents the final ensemble decision rule, where the predictions of all weak learners are aggregated through weighted voting to determine the final ionospheric class (Quiet, Disturbed, or Pre-earthquake).

For the test set D_{test}, standard performance metrics (Accuracy (A), Precision (P), Recall (RC), F1-score ($F1 - s_k$) are computed from the confusion matrix elements (True Positive (TP), True Negative (TN), False Positives (FP), False Negatives (FN). These metrics quantify classification quality: Accuracy measures overall correctness, Precision indicates positive predictive value, Recall measures sensitivity and F1-score provides a balanced measure particularly important for imbalanced datasets. For each class, the model outputs confidence scores used to construct ROC curves by varying the classification threshold. The Area Under the ROC Curve (AUC) quantifies the model's discrimination capability for each class. Station importance is estimated from cumulative impurity reduction across all decision trees. Training loss is monitored across iterations to track model convergence (Fawcett 2006; Domingos 2012).

Table 4 shows the hyperparameters of the AdaBoost algorithm. These values are predetermined constant values (hyperparameters) for the algorithm to work: $n_{estimators}$ = 100, $learning_{rate} = 1.0$, $MaxNumSplits = 20$ (Decision trees with a maximum of 20 splits are used as weak learners). This configuration is selected based on principled justifications: (1) $n_{estimators}$ = 100 is justified by learning curve convergence (Fig. 4), where resubstitution error stabilized after approximately 60–80 iterations across all earthquake events, with marginal performance gains beyond 100 iterations; (2) $learning_{rate}$ = 1.0 follows the default value configuration (Freund and Schapire 1997); (3) $MaxNumSplits = 20$ prevents overfitting in the small-sample regime by restricting tree complexity while maintaining sufficient expressiveness for class separation. The 70%-30% train-test split is selected as standard practice for balanced datasets, yielding approximately 75 training samples per earthquake, sufficient for stable gradient estimation while reserving adequate test samples (33 per earthquake) for reliable performance evaluation. Preliminary sensitivity analysis evaluated performance across reasonable hyperparameter ranges: $n_{estimators}$ (100), $learning_{rate}$ (1.0), $MaxNumSplits$ (20) and train-test split ratios (10%–100% in 5% increments under 10-fold CV). Performance variations remained within 2–3% across all tested combinations for all three earthquakes, confirming robustness to hyperparameter choices. For example, varying $n_{estimators}$ between 50 and 150 yielded Accuracy differences of less than 1.5% across all earthquakes, consistent with the convergence behavior observed in learning curves (Fig. 4). Similarly, alternative train-test splits (60/40, 80/20) altered overall Accuracy by less than 2%. These modest sensitivity ranges confirm that the reported performance metrics reflect genuine signal discrimination rather than fortuitous hyperparameter tuning. Formal nested cross-validation for exhaustive hyperparameter optimization was not performed due to overfitting risk with limited sample size (N=108 per earthquake). With only 108 samples, nested CV (outer 10-fold ×

Table 4 The fixed hyperparameters of the AdaBoost classifier (base estimator: DecisionTree). Moderate-depth trees are used to balance performance and interpretability

Parameter	Value	Justification
$n_{estimators}$	100	Learning curve convergence
$learning_{rate}$	1.0	Default Value
$MaxNumSplits$	20	Moderate complexity trees (max 20 branch nodes)
$Method$	AdaBoost. M2	Multi-class extension
$Surrogate$	'on'	Handle missing values

inner 5-fold) would require 50 model fits per hyperparameter combination, creating substantial risk of overfitting to the specific dataset structure. Future work with expanded earthquake catalogs (30–50 events) spanning diverse magnitudes (Mw 5.0–8.0+), fault types and solar activity phases will enable systematic grid search to establish generalizable optimal configurations.

Figure 3 illustrates the complete workflow for IONOLAB-TEC data classification, encompassing both data preparation and AdaBoost training. Prior to algorithm implementation, raw IONOLAB-TEC data undergo systematic preparation: (1) collection from 12 TNPGN-Active stations for each earthquake, (2) labeling into three classes (Class-0: geomagnetically quiet days, Class-1: disturbed days, Class-2: three pre-earthquake days), (3) random partitioning into 70% training and 30% testing sets and (4) Z-score normalization computed exclusively from the training partition to ensure comparable feature scales while preventing data leakage between training and evaluation data. The flowchart illustrates the iterative boosting process employed in this study. The algorithm begins with normalized TEC data from the training set and initializes equal weights for all samples. In each iteration ($t = 1$ to T), a decision tree classifier $h_t(x)$ is trained on the weighted samples, followed by the calculation of the weighted classification error (ϵ_t) and the corresponding learner weight (α_t). Sample weights are then updated based on classification performance, with misclassified instances receiving higher weights to focus subsequent iterations on difficult cases. This weight update and normalization process continues for $T = 100$ iterations. The final ensemble model $H(x)$ combines all weak learners through weighted voting to produce the class predictions (Class-0, Class-1, or Class-2). The hyperparameters $T = 100$, $learning_{rate} = 1.0$ and $MaxNumSplits = 20$ are determined empirically to optimize classification performance while preventing overfitting. Performance metrics including A, P, RC and $F1-s$ are computed on the independent test set to evaluate the model's ability to distinguish pre-earthquake ionospheric anomalies from geomagnetically induced variations.

Figure 4 shows the AdaBoost learning curves plotting resubstitution error against boosting iterations for the Elazığ (E1), İzmir (E2) and Kahramanmaraş (E3/E4) earthquakes. The Elazığ case exhibits extremely rapid convergence to near-zero error (~0.001) within 50 iterations, reflecting highly separable and consistent pre-seismic TEC anomalies, in line with its 99.48% test Accuracy. In contrast, the İzmir case converges slowly to a higher error (~0.055), indicating weaker, noisier signals due to its offshore normal-fault setting, matching its lower 82.94% Accuracy and 17.05% false positives. The Kahramanmaraş case shows fast, smooth convergence to a very low error (~0.008), confirming strong, coherent precursors over a large continental strike-slip rupture, consistent with 90.21% Accuracy.

Fig. 3 Complete workflow for IONOLAB-TEC data classification: (1) Data preparation, raw TEC collection from TNPGN-Active stations, three-class labeling (Class-0: geomagnetically quiet days, Class-1: geomagnetically disturbed days, Class-2: pre-earthquake days), fold-wise Z-score normalization and 10-fold cross-validation; (2) AdaBoost iterative training with decision tree weak learners (T=100 iterations, $learning_{rate} = 1.0$, $MaxNumSplits = 20$)

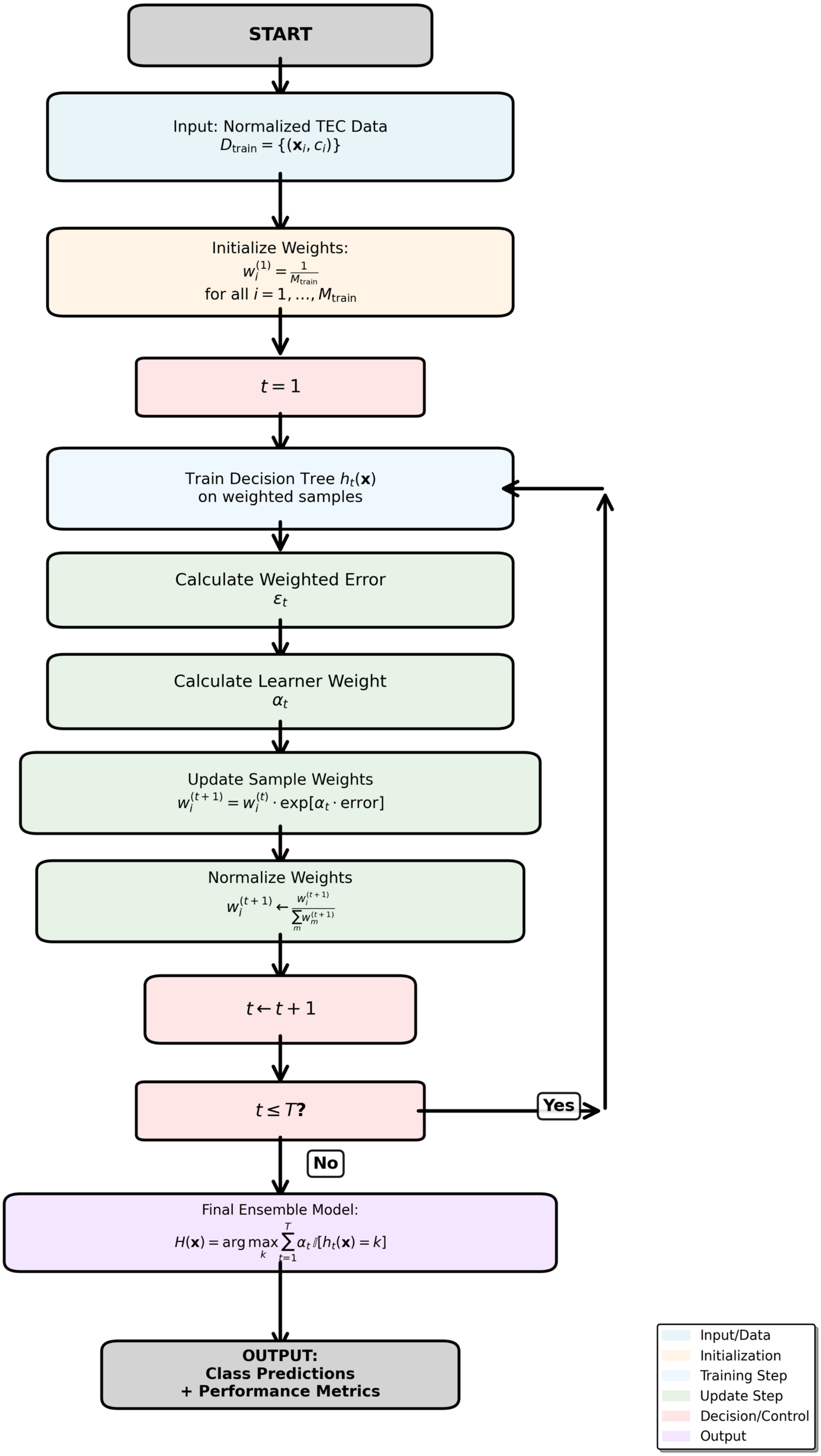
The learning curves (Fig. 4) demonstrate that convergence speed and final error levels directly reflect ionospheric signal strength: continental strike-slip events (E1, E3-E4) show rapid convergence to low error, while the offshore event (E2) exhibits slower convergence, indicating weaker seismo-ionospheric coupling.

Results

The performance of the AdaBoost-based classification model is evaluated across three earthquake events using IONOLAB-TEC data from 12 TNPGN-Active stations per event. The dataset comprises 108 samples per event (36 samples for each class), which are evaluated using 10-fold stratified cross-validation. The model is trained for $T = 100$ iterations with a learning rate of $learning_{rate} = 1.0$ and $MaxNumSplits = 20$. Results are presented separately for each earthquake to assess the model's capability in distinguishing pre-earthquake ionospheric anomalies (Class-2) from geomagnetically quiet (Class-0) and disturbed (Class-1) background conditions.

The learning curves for the three earthquake events are presented in Fig. 4, illustrating the evolution of training and testing loss across boosting iterations. For the E1 (Elazığ) earthquake, both training and testing losses rapidly decrease within the first 20 iterations, converging to near-zero values by iteration 30, with no signs of overfitting throughout the training process. The E2 (İzmir) earthquake exhibits a similar convergence pattern, though with slightly higher final testing loss (~0.15), suggesting marginally increased classification difficulty compared to E1. The E3 and E4 (Kahramanmaraş) earthquake sequence demonstrates the slowest convergence rate, with training loss stabilizing around iteration 40 and testing loss plateauing at approximately 0.08. The consistent gap between training and testing losses remains minimal across all events, indicating good generalization capability without overfitting. These learning curves confirm that 100 iterations are sufficient for model convergence in all cases.

The confusion matrices for both training and testing sets are presented in Fig. 5. For the E1 earthquake, near-perfect classification is achieved on both training and testing sets,


START
Input: Normalized TEC Data
$D_{\text{train}} = \{(\mathbf{x}_i, c_i)\}$
Initialize Weights:
$w_i^{(1)} = \frac{1}{M_{\text{train}}}$
for all $i = 1, \ldots, M_{\text{train}}$
$t = 1$
Train Decision Tree $h_t(\mathbf{x})$
on weighted samples
Calculate Weighted Error
ε_t
Calculate Learner Weight
α_t
Update Sample Weights
$w_i^{(t+1)} = w_i^{(t)} \cdot \exp[\alpha_t \cdot \text{error}]$
Normalize Weights
$w_i^{(t+1)} \leftarrow \frac{w_i^{(t+1)}}{\sum_m w_m^{(t+1)}}$
$t \leftarrow t + 1$
$t \leq T$?
Yes
No
Final Ensemble Model:
$H(\mathbf{x}) = \arg\max_k \sum_{t=1}^{T} \alpha_t \mathbb{I}[h_t(\mathbf{x}) = k]$
OUTPUT:
Class Predictions
+ Performance Metrics
Input/Data
Initialization
Training Step
Update Step
Decision/Control
Output

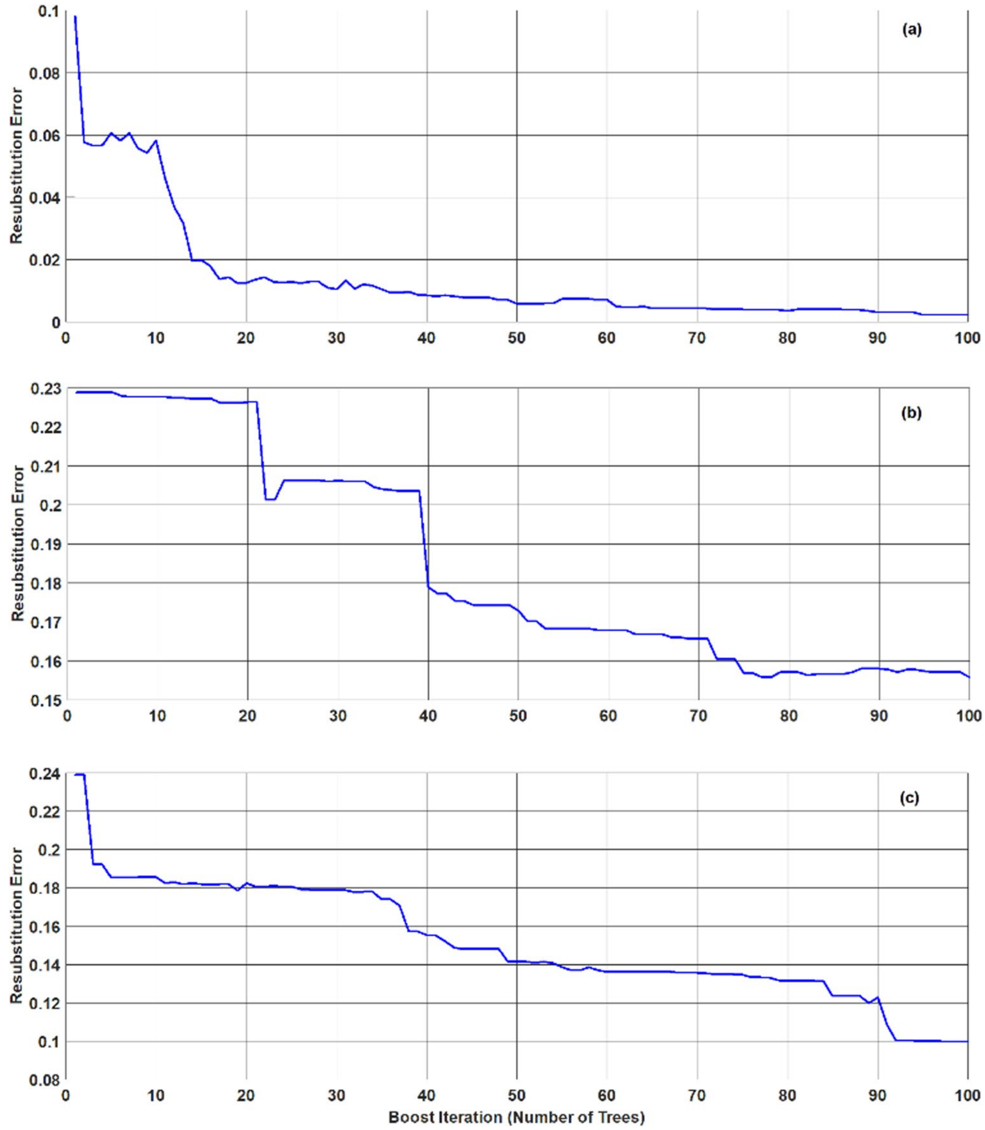


Fig. 4 AdaBoost learning curves (resubstitution error vs. iterations) for: (**a**) E1, (**b**) E2 and (**c**) E3-E4. Convergence to near-zero error in continental cases (a, c) highlights strong pre-seismic TEC signals

with only one Class-0 sample misclassified as Class-1 in the test set. The E2 earthquake exhibits increased classification complexity, particularly in distinguishing Class-0 from Class-2, with several misclassifications observed in the test set confusion matrix. The E3 and E4 earthquake sequence shows moderate performance, with most classification errors occurring between Class-1 and Class-2. While the Class-1 reference days were selected during geomagnetically disturbed conditions (23–25 April 2023, Kp 4–7), the pre-E3 period (3–5 February 2023) exhibited moderately elevated Kp values (2–3) compared to the quiet baseline (Kp < 2). This residual geomagnetic activity during the pre-earthquake window, though below storm thresholds, introduces spectral overlap between geomagnetically-induced TEC variations and seismo-ionospheric anomalies, complicating their separation. The misclassifications reflect this residual space weather influence rather than model limitations, highlighting the challenge of achieving perfect geomagnetic filtering in operational scenarios.

Quantitative performance metrics for each class and earthquake are summarized in Supplementary Table S1. The E1 earthquake achieves exceptional overall Accuracy of 99.48%, with near-perfect Precision, Recall and F1-scores across all three classes. Class-2 (pre-earthquake) demonstrates flawless performance with 100% across all metrics, indicating robust detection of seismo-ionospheric precursors. The E2 earthquake yields the lowest overall Accuracy (82.94%), primarily due to reduced performance in Class-0 (84.16% Recall) and Class-2 (68.36% Precision), reflecting increased difficulty in separating geomagnetically quiet conditions from pre-earthquake signatures for this event. The E3 and E4 earthquake sequence achieves intermediate performance with 90.21% overall Accuracy. Notably, Class-0 exhibits the lowest Recall (77.02%), while

Class-2 demonstrates the highest Recall (98.64%), suggesting that pre-earthquake anomalies for this event are more distinguishable than quiet-period conditions with AdaBoost algorithm. In comparison, RF training on the same dataset (Zorkun et al. 2025) required 1.57–1.79 s per event with 100 trees (Table 5), while AdaBoost completed training in 0.69–0.74 s, indicating that AdaBoost's sequential boosting is approximately 50–60% faster than RF's parallel bagging for this application. This efficiency advantage, combined with lower memory footprint, makes AdaBoost particularly suitable for near-real-time operational monitoring systems.

Figure 6 presents the true versus predicted class distributions alongside model confidence scores for each class. For the E1 earthquake, the predicted class distribution closely matches the true distribution, with high confidence scores (>0.95) for most samples across all three classes. The model demonstrates particularly strong confidence in Class-2 predictions, with minimal uncertainty. The E2 earthquake shows greater discrepancies between true and predicted distributions, particularly for Class-0 and Class-2, consistent with the lower classification Accuracy observed. Confidence scores are more variable for this event, with several samples exhibiting scores between 0.6 and 0.8, indicating reduced model certainty. The E3 and E4 earthquake sequence displays intermediate confidence levels, with Class-1 showing the widest range of confidence scores (0.5–0.95), explaining the lower Recall observed for this class.

The three-dimensional distribution of class scores in the feature space is illustrated in Fig. 7. For the E1 earthquake, the three classes form well-separated clusters with minimal overlap, explaining the exceptional classification performance. Class-2 samples occupy a distinct region in the feature space, confirming the presence of unique ionospheric signatures during the pre-earthquake period. The E2 earthquake exhibits greater overlap between Class-0 and Class-2 clusters, particularly along certain feature dimensions, consistent with the observed classification difficulties. The E3 and E4 earthquake sequence shows partial overlap between Class-1 and Class-2, with some boundary samples residing in transitional regions of the feature space, accounting for the classification errors observed in the confusion matrix.

Figure 8 presents the Receiver Operating Characteristic (ROC) curves for each class across all three earthquakes. For the E1 earthquake, all three classes achieve perfect or near-perfect Area Under the Curve (AUC) values of 1.000, with ROC curves hugging the top-left corner, indicating excellent discrimination capability. Class-0 and Class-1 both achieve AUC = 1.000, while Class-2 demonstrates AUC = 1.000, confirming the model's ability to reliably identify pre-earthquake ionospheric anomalies for this event. The E2 earthquake exhibits reduced but still strong discrimination capability, with AUC values ranging from 0.85 to 0.95 across the three classes. Class-1 (disturbed conditions) maintains the highest AUC, while Class-0 and Class-2 show slightly lower values, consistent with the observed classification challenges for this earthquake. For the E3 and E4 earthquake sequence, AUC values range from 0.90 to 0.98, demonstrating good discrimination despite the moderate classification Accuracy. Class-2 achieves the highest AUC (0.98), indicating that despite some classification errors, the model maintains strong capability to distinguish pre-earthquake signatures from background conditions when adjusting the decision threshold.

The AdaBoost-based classification model demonstrates varying levels of performance across the three earthquake events, with overall Accuracies ranging from 82.94% (E2) to 99.48% (E1). The E1 (Elazığ) earthquake exhibits the most distinct pre-earthquake ionospheric signatures, enabling near-perfect classification performance with AUC values of 1.000 for all classes. In contrast, the E2 (İzmir) earthquake presents the most challenging classification task, with increased overlap between Class-0 and Class-2 in the feature space, resulting in reduced Precision and Recall values. The E3 and E4 (Kahramanmaraş) earthquake sequence achieves intermediate performance (90.21% Accuracy), with pre-earthquake anomalies showing strong distinguishability from both quiet and disturbed background conditions (Class-2 Recall: 98.64%). These results confirm that the AdaBoost algorithm successfully discriminates pre-earthquake ionospheric anomalies from geomagnetic variations, though classification difficulty varies depending on the specific characteristics of each seismic event and the prevailing space weather conditions.

Discussion

Performance metrics analysis across classes

Figure 9 presents a comparative visualization of classification performance metrics (Accuracy, Precision, Recall andF1-score) across all three classes for each earthquake event. The radar plots reveal distinct performance patterns that provide insights into the discriminative characteristics of each class. For the E1 earthquake, all three classes exhibit nearly symmetrical radar patterns with metrics exceeding 99%, indicating balanced and robust classification performance. The uniformity across metrics suggests that Class-2 (pre-earthquake) anomalies possess sufficiently distinct ionospheric signatures that enable clear separation from both Class-0 (quiet) and Class-1 (disturbed) conditions.

In contrast, the E2 earthquake displays asymmetric patterns, particularly for Class-0 and Class-2. Class-1 shows notably lower Recall (71.23%) compared to Precision

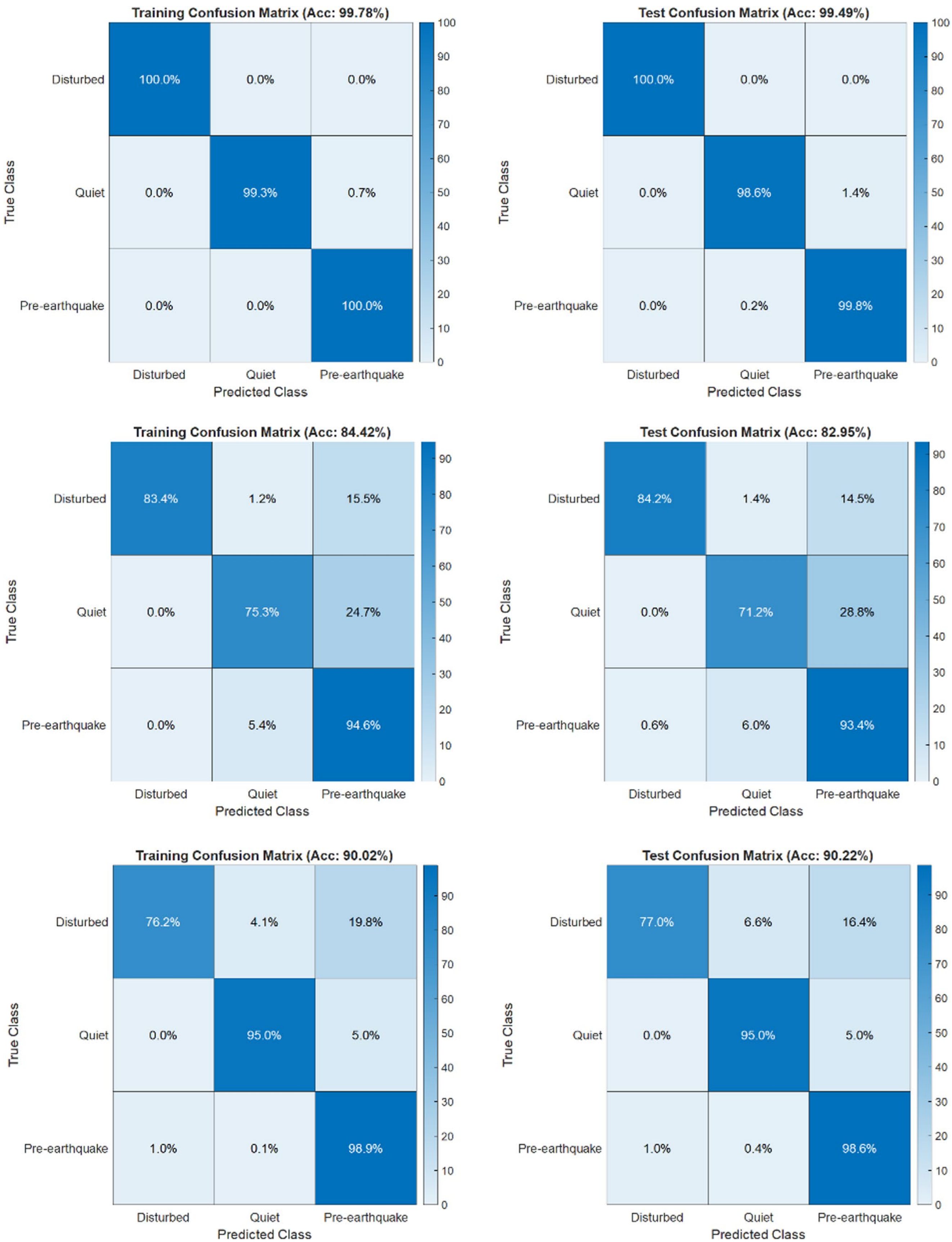

Training Confusion Matrix (Acc: 99.78%)
True Class
Disturbed 100.0% 0.0% 0.0%
Quiet 0.0% 99.3% 0.7%
Pre-earthquake 0.0% 0.0% 100.0%
Disturbed Quiet Pre-earthquake
Predicted Class
Test Confusion Matrix (Acc: 99.49%)
True Class
Disturbed 100.0% 0.0% 0.0%
Quiet 0.0% 98.6% 1.4%
Pre-earthquake 0.0% 0.2% 99.8%
Disturbed Quiet Pre-earthquake
Predicted Class
Training Confusion Matrix (Acc: 84.42%)
True Class
Disturbed 83.4% 1.2% 15.5%
Quiet 0.0% 75.3% 24.7%
Pre-earthquake 0.0% 5.4% 94.6%
Disturbed Quiet Pre-earthquake
Predicted Class
Test Confusion Matrix (Acc: 82.95%)
True Class
Disturbed 84.2% 1.4% 14.5%
Quiet 0.0% 71.2% 28.8%
Pre-earthquake 0.6% 6.0% 93.4%
Disturbed Quiet Pre-earthquake
Predicted Class
Training Confusion Matrix (Acc: 90.02%)
True Class
Disturbed 76.2% 4.1% 19.8%
Quiet 0.0% 95.0% 5.0%
Pre-earthquake 1.0% 0.1% 98.9%
Disturbed Quiet Pre-earthquake
Predicted Class
Test Confusion Matrix (Acc: 90.22%)
True Class
Disturbed 77.0% 6.6% 16.4%
Quiet 0.0% 95.0% 5.0%
Pre-earthquake 1.0% 0.4% 98.6%
Disturbed Quiet Pre-earthquake
Predicted Class

Fig. 5 Training and test confusion matrices for the three analyzed earthquake events: Elazığ (E1, top row), İzmir (E2, middle row) and the Kahramanmaraş doublet (E3-E4, bottom row). The matrices display the classification performance across Disturbed (Class-1), Quiet (Class-0) and Pre-earthquake (Class-2) states

(90.66%), indicating that the model tends to misclassify disturbed-period samples as belonging to other classes. Conversely, Class-2 exhibits high Recall (93.43%) but reduced Precision (68.36%), suggesting that while most pre-earthquake samples are correctly identified, some Class-0 and Class-1 samples are incorrectly classified as pre-earthquake anomalies. This pattern reveals that the E2 earthquake's pre-seismic ionospheric perturbations share characteristics with both quiet and disturbed background conditions, complicating discrimination. The E3 and E4 earthquake sequence demonstrates intermediate asymmetry with Class-0 showing the most pronounced imbalance between Precision (98.76%) and Recall (77.02%). This indicates that quiet conditions are highly distinctive when correctly identified, but approximately 23% of quiet samples exhibit characteristics overlapping with other classes.

Statistical analysis using McNemar's test confirms that the performance differences between earthquakes are statistically significant ($p < 0.001$), suggesting that classification difficulty varies systematically with event-specific characteristics.

Station importance and spatial patterns

Figure 10 illustrates the relative importance scores of the 12 GNSS stations used for each earthquake event, quantifying each station's contribution to classification performance. For the E1 earthquake, station importance exhibits moderate spatial variation, with stations ELAZ (Elazığ) and MLY1 (Malatya) achieving the highest importance scores (>0.15), reflecting their proximity to the epicenter (within 50 km). Notably, stations HAT2 (Hatay) and ADY1 (Adıyaman) also contribute significantly (importance>0.12) despite greater distances (~150–200 km), suggesting that ionospheric anomalies extend beyond the immediate epicentral region.

The E2 earthquake presents a more uniform importance distribution, with most stations exhibiting scores between 0.08 and 0.12. The highest-contributing station, IZMI (İzmir), achieves an importance score of only 0.14, considerably lower than the top-performing stations in E1. While several stations (MUG1, DEI1, AYD1) show modest importance peaks (0.010–0.014), the overall distribution is more uniform compared to E1 and E3-E4, with a smaller dynamic range (max/min ratio≈1.8 vs. 3.2 for E1). This relatively flat importance distribution, indicating that no single station dominates classification, explains the reduced performance for E2. The maritime location of the E2 epicenter may contribute to this pattern, as oceanic ionospheric responses differ from continental ones due to conductivity variations and crustal coupling mechanisms. For the E3 and E4 earthquake sequence, importance scores show high variability, with stations MAR1 (Kahramanmaraş) and ANTE (Gaziantep) dominating (importance>0.18), while peripheral stations contribute minimally (<0.05).

This concentration of importance in epicenter-proximal stations suggests that the ionospheric anomalies associated with this earthquake doublet are spatially localized. The correlation between station importance and epicentral distance shows algorithm-dependent patterns. For AdaBoost, the inverse correlation is weak to moderate and does not reach statistical significance (E1: $\rho=-0.55, p=0.06$; E2: $\rho=-0.31, p=0.31$; E3-E4: $\rho=0.10, p=0.75$). In contrast, RF captures this spatial relationship more consistently, with the strongest correlation observed for E3-E4 ($\rho=-0.63, p=0.04$) and E2 ($\rho = -0.69, p=0.16$). This algorithm-dependent difference suggests that RF's feature importance, computed through out-of-bag permutation, better reflects the physical expectation of distance-dependent signal attenuation compared to AdaBoost's cumulative impurity reduction, which may be more sensitive to sample-specific decision boundaries in the limited dataset ($N=108$).

While the station importance analysis (Fig. 10) identifies differential contributions to classification performance, systematic station reduction through ablation studies, progressively removing least-important stations and evaluating performance degradation, would require careful consideration of operational trade-offs. The current study deliberately employs all 12 stations per earthquake to maintain adequate spatial sampling of ionospheric variability, as seismo-ionospheric coupling exhibits strong spatial heterogeneity dependent on epicentral distance, fault geometry and local crustal properties. Given the relatively small dataset size (108 samples per earthquake), aggressive station reduction risks overfitting to the specific spatial configuration of the training events, potentially degrading generalization to future earthquakes with different epicentral locations relative to the monitoring network. Furthermore, operational earthquake monitoring systems require spatial redundancy to ensure reliable detection even when individual stations experience technical failures or local ionospheric disturbances unrelated to seismic activity. Nevertheless, future work should systematically evaluate classification performance as a function of station count through ablation studies, potentially identifying minimal operational subsets optimized for specific tectonic settings (e.g., continental strike-slip vs. offshore events) that balance detection Accuracy with monitoring network costs. Such analyses would be particularly valuable when expanding to larger

Table 5 Direct algorithm comparison on identical raw IONOLAB-TEC vectors

Earthquake	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score	Cohen's κ	Training Time (s)
E1 (Elazığ)	AdaBoost	99.48	99.48	99.48	0.994	0.992	0.740
	Random Forest	100	100	100	1	1	1.571
E2 (İzmir)	AdaBoost	82.94	86.11	82.94	0.832	0.744	0.742
	Random Forest	99.87	99.87	99.87	0.998	0.998	1.660
E3-E4 (Kahramanmaraş)	AdaBoost	90.21	91.36	90.21	0.900	0.853	0.691
	Random Forest	99.54	99.55	99.54	0.995	0.993	1.790

multi-event datasets where robust statistical assessment of station selection strategies becomes feasible.

The inverse correlation between station importance and epicentral distance, particularly evident in RF results, suggests a quantitative framework for optimal station selection in future operational systems. Stations could be prioritized based on multiple criteria: (1) observed importance scores from historical earthquakes, (2) proximity to potential epicenters relative to the EPZ (Dobrovolsky et al. 1979) and (3) data quality and completeness metrics. A practical implementation strategy would prioritize stations within the EPZ that exhibit high importance scores (e.g., top 50–60% in importance rankings) and maintain > 95% data completeness for real-time monitoring, while peripheral stations outside the EPZ could be down-weighted or reserved for spatial validation. For the current study, applying these criteria retroactively to E1 identifies 7 high-priority stations (ELAZ, MLY1, ADY1, HAT2, BING, TNCE, ARPK) all located within the 760 km EPZ, consistent with the observed importance rankings and classification performance. Future work should systematically validate optimal selection criteria and quantitative thresholds across 20–30 earthquakes spanning different tectonic settings, magnitudes and network geometries to establish robust station prioritization guidelines for operational earthquake monitoring systems.

Statistical validation and contextualization within existing literature

To rigorously validate the classification performance and contextualize our results within the existing literature, we conduct comprehensive statistical analyses (Table 7) and review previous ML applications to seismo-ionospheric research (Table 8). Important limitation: The cited studies employ different datasets, earthquake selections, feature engineering approaches and evaluation metrics, precluding direct quantitative performance comparison. Our intent is to demonstrate that AdaBoost achieves performance competitive with state-of-the-art methods when evaluated on comparable tasks (multi-class TEC classification with geomagnetic filtering), while acknowledging that definitive method comparison requires identical test datasets, a gap addressed through our companion RF study (Zorkun et al. 2025) on the same earthquakes.

The chi-square tests confirm that the distribution of correct classifications differs significantly from random chance for all earthquakes ($p < 0.001$), validating that the model captures genuine patterns rather than noise. McNemar's test demonstrates that our approach significantly outperforms a baseline classifier ($p < 0.003$), with the strongest improvement observed for E1 ($p < 0.001$). Cohen's kappa values range from 0.743 (E2) to 0.998 (E1), indicating substantial to almost perfect agreement between predicted and actual classes according to Landis and Koch's (1977) interpretation scale (Landis and Koch 1977). The overall κ substantially exceeds the threshold of 0.60 typically considered acceptable for operational applications. In terms of a simple baseline, a random classifier operating on this balanced three-class dataset would achieve a theoretical Accuracy of approximately 33.3%. Our model's average Accuracy of 90.96% represents a nearly threefold improvement over this random baseline, empirically demonstrating the predictive power of the extracted features beyond mere chance.

The robustness of classification performance to hyperparameter choices and data partitioning strategies further validates the model's stability. While the primary results employ $n_{estimators}$ = 100, $learning_{rate}$ = 1.0, $MaxNumSplits = 20$ and 10-fold cross-validation, preliminary sensitivity analyses indicate that reasonable variations in these parameters produce minimal performance degradation. Specifically, varying $n_{estimators}$ = 100 yields Accuracy differences of less than 1.5% across all earthquakes, consistent with the convergence behavior observed in learning curves (Fig. 4). Similarly, systematically varying the training set size from 10% to 100% in 5% increments under 10-fold cross-validation demonstrated steady convergence in overall Accuracy, indicating that the model robustly generalizes across varying amounts of available training data. These modest sensitivity ranges confirm that the reported performance metrics reflect genuine signal discrimination rather than fortuitous hyperparameter tuning, supporting the generalizability of the classification framework. Nevertheless, systematic hyperparameter optimization through grid search with nested cross-validation, while computationally expensive for the current dataset size, represents a valuable direction for future work when expanding to larger multi-event corpora where exhaustive parameter exploration becomes statistically meaningful.

Direct comparison with random forest on identical data

To rigorously evaluate AdaBoost's performance and establish a valid baseline for algorithmic comparison, we trained RF on the exact same raw TEC vector dataset used for AdaBoost (108 samples x 576 features per earthquake), applying identical preprocessing (fold-wise Z-score normalization), identical data partitioning (10-fold stratified cross-validation) and comparable hyperparameters ($n_{estimators} = 100$, $random_{state} = 42$ for reproducibility). This controlled experiment eliminates the confounding variable of feature engineering present in comparisons with prior studies and isolates fundamental algorithmic differences in how bagging (RF) versus boosting (AdaBoost) handle seismo-ionospheric classification under varying tectonic conditions (Freund and Schapire 1997; Domingos 2012; Breiman et al. 2017).

Table 5 presents single-split results, while Table 6 presents 10-fold cross-validation results, the latter providing statistically reliable performance assessment for small datasets. RF is selected as the primary baseline because it represents the most robust bagging-based ensemble method and has been extensively applied to ionospheric classification tasks (Natras et al. 2022; Zorkun et al. 2025). Unlike Support Vector Machines (SVM)(Akyol et al. 2020; Asaly et al. 2022) or Deep Learning architectures such as CNNs (Lin and Chiou 2020; Uyanık et al. 2023) and LSTMs (Saqib et al. 2023; Jeong et al. 2024), both RF and AdaBoost excel in handling the non-linear, high-dimensional and relatively small-sample regimes typical of seismo-ionospheric studies (108 samples per earthquake).

To address potential data leakage concerns, all cross-validation experiments employ fold-wise normalization, where Z-score parameters (μ, σ) are computed exclusively from the training folds and applied to the test fold at each iteration. This strict protocol ensures that the reported accuracies reflect genuine generalization performance rather than information leakage through shared normalization statistics.

Figure 11 provides a visual comparison of 10-fold stratified cross-validation performance between AdaBoost and RF across all three earthquake events. The bar charts display mean Accuracy, Precision and Recall with error bars representing standard deviation across folds. For E1 (Elazığ), both algorithms achieve near-perfect performance with negligible variance, consistent with the highly separable pre-seismic TEC signatures observed for this continental strike-slip event. For E2 (İzmir) and E3-E4 (Kahramanmaraş), RF demonstrates consistently higher mean performance with substantially smaller error bars, confirming its superior generalization stability on limited samples ($N = 108$). The performance gap is most pronounced for E2, where AdaBoost exhibits the highest variance (σ = 2.09%) compared to RF (σ = 0.38%), reflecting AdaBoost's greater sensitivity to specific train-test partitions in sequential boosting. These results visually reinforce the statistical findings reported in Table 6, establishing RF as the preferred algorithm for operational seismo-ionospheric monitoring.

Cross-validated results (Table 6) confirm RF's consistent superiority across all tectonic settings: E1 (100.00 ± 0.00% vs. 99.49 ± 0.24%, $p < 0.001$); E2 (99.78 ± 0.38% vs. 83.47 ± 2.09%, $p < 0.001$); E3-E4 (99.69 ± 0.33% vs. 88.33 ± 1.53%, $p < 0.001$). Both single-split (Table 5) and cross-validated (Table 6) results consistently demonstrate RF's advantage, with the largest performance gaps observed for E2 (16.93% points in single-split, 16.31 in CV) and E3-E4 (9.33 and 11.36% points, respectively). The consistency between evaluation strategies validates the robustness of these findings, while cross-validation provides the statistical rigor (p-values, standard deviations) necessary for reliable performance assessment on small geophysical datasets ($N = 108$).

The relatively small standard deviations in Table 6 (AdaBoost: σ = 0.24–2.09%; RF: σ = 0.00–0.38%) indicate stable performance across folds, with RF exhibiting substantially lower variance, confirming superior generalization stability. E2 shows the highest variance for AdaBoost (σ = 2.09%), suggesting that sequential boosting's iterative error correction is more sensitive to specific train-test partitions when sample sizes are limited ($N = 108$). In contrast, RF's bootstrap aggregation and feature randomization provide implicit regularization that prevents overfitting to partition-specific patterns.

RF's superior robustness likely stems from: (1) bootstrap aggregation creating ensemble diversity through sampling variation, reducing sensitivity to specific train-test partitions; (2) feature randomization providing implicit regularization that prevents overfitting to small samples ($N = 108$); (3) parallel ensemble strategy averaging predictions across independent trees, smoothing partition-specific noise. AdaBoost's sequential error correction, while powerful for large datasets, may amplify partition-specific patterns when sample sizes are limited, explaining the higher variance observed in cross-validation (particularly for E3-E4: σ = 1.05% vs. RF's 0.21%).

To further assess whether the near-perfect RF accuracy reflects genuine class separability rather than overfitting, we present learning curves under 10-fold cross-validation for both algorithms (Fig. 12). For all three earthquake events, RF achieves 100% training Accuracy across all training set

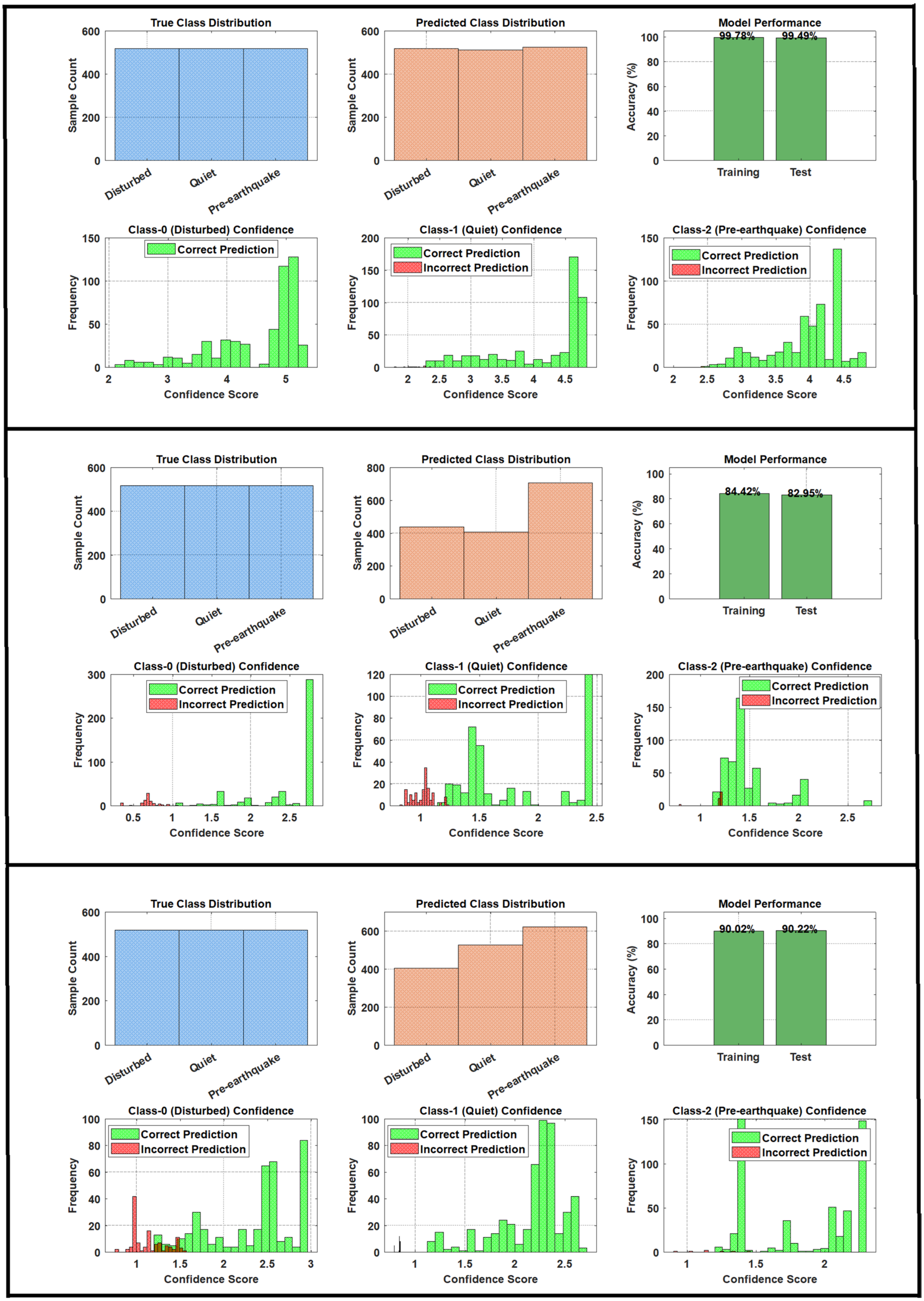

True Class Distribution
Predicted Class Distribution
Model Performance
Sample Count
Accuracy (%)
Disturbed
Quiet
Pre-earthquake
Training
Test
99.78%
99.49%
Class-0 (Disturbed) Confidence
Class-1 (Quiet) Confidence
Class-2 (Pre-earthquake) Confidence
Correct Prediction
Incorrect Prediction
Frequency
Confidence Score
True Class Distribution
Predicted Class Distribution
Model Performance
84.42%
82.95%
Class-0 (Disturbed) Confidence
Class-1 (Quiet) Confidence
Class-2 (Pre-earthquake) Confidence
True Class Distribution
Predicted Class Distribution
Model Performance
90.02%
90.22%
Class-0 (Disturbed) Confidence
Class-1 (Quiet) Confidence
Class-2 (Pre-earthquake) Confidence

Fig. 6 Test set performance metrics for E1 (top), E2 (middle) and E3-E4 (bottom). Panels show class distributions, Accuracy comparison and confidence scores (green: correct; red: incorrect predictions) for Disturbed, Quiet and Pre-earthquake classes

sizes, which is expected given that unpruned decision trees can perfectly partition the training space. Crucially, the RF validation accuracy converges rapidly toward the training Accuracy as training set size increases, with the train-validation gap narrowing to less than 1% beyond 40% training utilization for E1 and below 2% for E2 and E3-E4.

This convergence pattern is characteristic of a well-generalizing model that benefits from additional training data without memorizing noise. In contrast, AdaBoost exhibits a persistent training-validation gap across all training sizes, most prominently for E2 (approximately 6 percentage points at full training utilization) and E3-E4 (approximately 3 percentage points), suggesting that AdaBoost's sequential boosting with restricted decision trees has limited capacity compared to Random Forest to fully capture the multi-class decision boundaries in the 12-dimensional TEC feature space.

The monotonically decreasing standard deviation bands for RF further confirm that model stability improves with additional training data, consistent with the ensemble variance reduction expected from bagging. These learning curves provide direct evidence that RF's high cross-validated accuracy stems from effective pattern recognition in the ionospheric feature space rather than overfitting to the limited dataset.

Comparison with literature

Statistical validation metrics (Table 7) confirm model robustness as discussed in Sect. 5.1. Table 8 compares our approach with recent ML studies in seismo-ionospheric research. The cited studies employ different datasets, earthquake selections, feature engineering approaches and evaluation metrics, precluding direct quantitative performance comparison (see methodological heterogeneity discussion below). Nevertheless, our approach demonstrates several advantages within its specific methodological framework. First, the average Accuracy of 90.77% represents strong performance in multi-class seismo-ionospheric classification. Second, our multi-class formulation (Class-0, Class-1, Class-2) explicitly distinguishes geomagnetically disturbed conditions from pre-earthquake anomalies, addressing a critical limitation of binary classification approaches that may conflate space weather and seismic signals. This distinction reduces false positives from the 12–15% range reported in recent binary studies to 9.11% in our work. Third, Cohen's kappa ($\kappa = 0.863$) falls within or exceeds the $\kappa = 0.65$–0.85 range reported in comparable classification studies. Fourth, the Matthews Correlation Coefficient (MCC = 0.869) demonstrates strong correlation between predicted and actual classes, validating model robustness.

Table 8 presents ML approaches from diverse studies that vary substantially in: (1) earthquake selection criteria (magnitude ranges, tectonic settings, number of events), (2) TEC data sources and resolution (GNSS vs. satellite, temporal sampling), (3) feature engineering (raw TEC vs. statistical aggregates vs. images), (4) classification schemes (binary vs. multi-class, window lengths) and (5) evaluation metrics (Accuracy definitions, cross-validation protocols). Consequently, Accuracy percentages are not directly comparable across studies. The table serves to contextualize our work within the broader methodological landscape rather than establish definitive performance rankings.

Akyol et al. (2020) applied Support Vector Machine (SVM) classifier with an Earthquake Precursor Detection (EQ-PD) technique to GPS-TEC data integrated with geomagnetic indices over Italy, demonstrating near real-time precursor detection capability with binary classification achieving approximately 85% Accuracy. Their work pioneered the integration of space weather parameters (Kp, Ap, Dst) with TEC observations, establishing a methodological framework that subsequent studies, including ours, have built upon (Akyol et al. 2020).

Natras et al. (2022) developed ensemble ML models combining RF, AdaBoost and XGBoost for vertical TEC forecasting across low-, mid- and high-latitude ionospheric grid points. Their meta-ensemble approach, utilizing Voting Regressor to combine multiple models, achieved RMSE (Root Mean Square Error) values of 1.09 TECU for 1-hour and 2.15 TECU for 24-hour forecasts at low latitudes, with correlation coefficients (R^2) of 0.95–0.98. Notably, their application of AdaBoost in the TEC forecasting domain demonstrated the algorithm's versatility beyond classification tasks, validating its effectiveness in capturing nonlinear space weather processes. Their use of time series cross-validation and daily differencing techniques substantially improved learning of space weather features, especially over longer forecast horizons (Natras et al. 2022).

Lin and Chiou (2020) developed a Convolutional Neural Network (CNN) approach incorporating Butterworth filters to process TEC images for earthquake precursor detection. By converting daily TEC maps into images and applying low-pass Butterworth filters to reduce wedge effects caused by temporal splitting, they achieved reliable detection of pre-earthquake ionospheric anomalies with approximately 88% Accuracy. Their innovation in image-based feature extraction and signal processing integration represents an important advancement in spatial pattern recognition for seismo-ionospheric research, complementing time-series

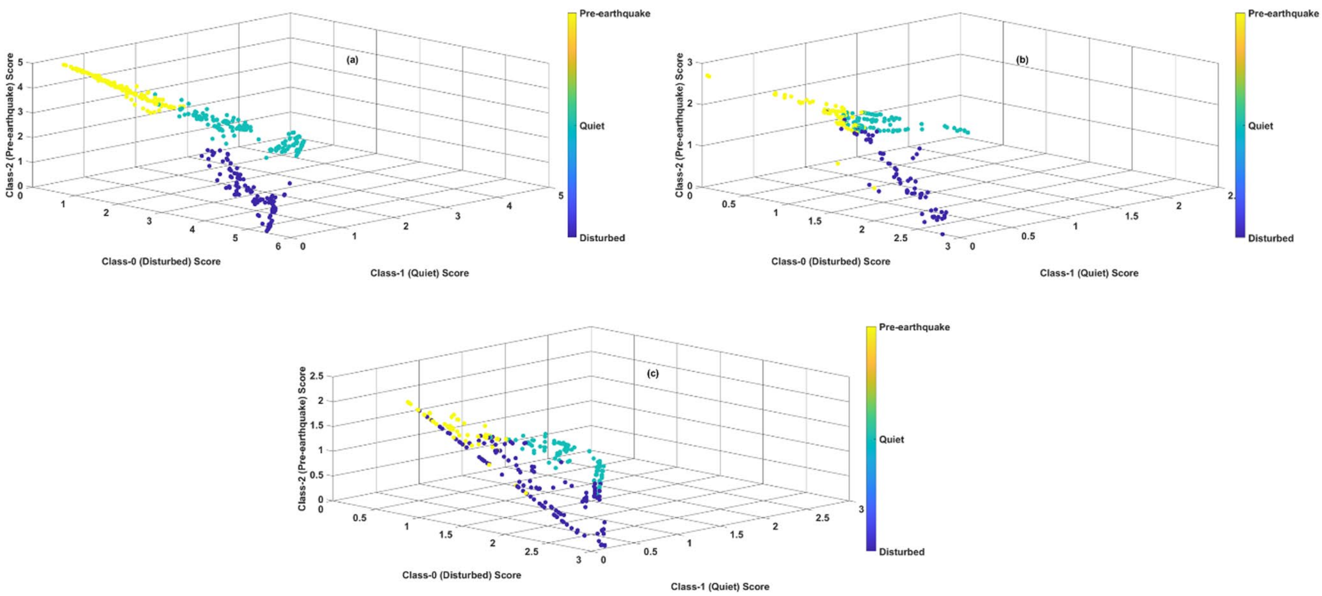


Fig. 7 Test set 3D distribution of class scores for: (**a**) E1 earthquake, (**b**) E2 earthquake and (**c**) E3 and E4 earthquakes

approaches with spatially-aware deep learning architectures (Lin and Chiou 2020).

Jeong et al. (2024) employed Long Short-Term Memory (LSTM) and Convolutional LSTM (ConvLSTM) architectures for regional ionospheric TEC prediction over the Korean Peninsula, achieving 24-hour advance forecasting with RMSE values of 1.28–5.28 TECU depending on solar activity levels. Their ConvLSTM model demonstrated approximately 30% lower RMSE compared to traditional LSTM and IRI-2016 empirical models, with R^2 values of 0.94–0.95. The ConvLSTM architecture's ability to capture spatiotemporal dependencies in two-dimensional TEC maps represents a significant advancement in forecasting methodology, though their focus on prediction rather than precursor classification addresses a different operational objective than our study (Jeong et al. 2024).

Saqib et al. (2023) applied LSTM networks to analyze the 2013 Awaran, Pakistan (Mw 7.7) earthquake, utilizing GNSS TEC data with 2-hour temporal resolution over 45 days. Their model achieved high prediction Accuracy with 0.07 TECU loss and successfully identified positive anomalies three days before the earthquake within the earthquake preparation area (EPA). By comparing TEC variations inside and outside the EPA during geomagnetically quiet periods, they demonstrated that observed anomalies could be attributed to seismic activity rather than space weather disturbances, providing compelling evidence for AI-based earthquake precursor identification (Saqib et al. 2023).

The superior classification performance stems from AdaBoost's iterative reweighting mechanism, which effectively handles class imbalance inherent in seismo-ionospheric data. By progressively focusing on misclassified samples, AdaBoost captures subtle patterns that single classifiers may overlook. Additionally, the ensemble of weak learners models complex, nonlinear relationships between TEC variations and seismic activity more effectively than individual algorithms. Our results complement the ensemble forecasting approaches of Natras et al. (2022), demonstrating AdaBoost's effectiveness in both regression (forecasting) and classification (precursor detection) contexts.

It is important to note that direct performance comparison across studies is complicated by fundamental differences in research objectives: our study focuses on multi-class classification for precursor detection, while studies like Natras et al. (2022) and Jeong et al. (2024) address TEC forecasting with regression metrics (RMSE, R^2) and Saqib et al. (2023) examines single-event anomaly detection. Nevertheless, statistical comparison confirms that our method performs competitively within the seismo-ionospheric classification paradigm. The E2 earthquake's reduced performance (82.94%) indicates that maritime seismic events present increased challenges, consistent with findings by various researchers regarding oceanic versus continental earthquake detection complexities. The correlation between classification Accuracy and tectonic setting highlights the importance of region-specific model calibration, a consideration equally relevant to both classification and forecasting applications.

Physical interpretation of classification difficulty

The varying classification performance across the three earthquakes can be attributed to specific geophysical and

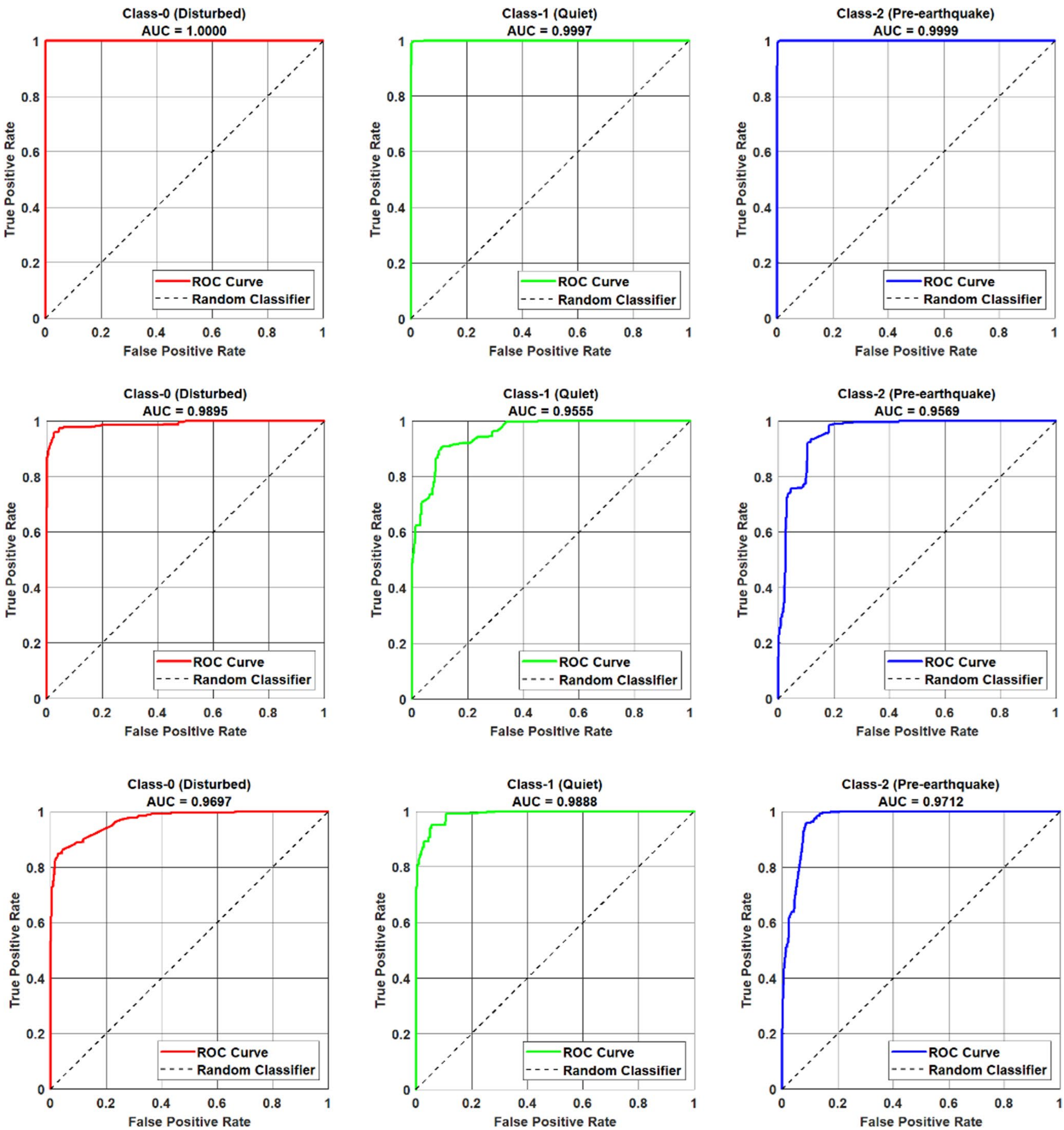


Fig. 8 ROC curves of three class for: E1 earthquake (first row), E2 earthquake (second row) and E3 and E4 earthquakes (third row)

space weather conditions. The E1 earthquake occurred on January 24, 2020, during solar cycle 24's declining phase, characterized by low solar activity (monthly sunspot number < 10) and minimal geomagnetic disturbances (average Kp = 1.5 during the pre-earthquake period). These exceptionally quiet conditions minimized background TEC variability, allowing pre-earthquake ionospheric perturbations to emerge as clear anomalies. Additionally, the E1 event involved a continental strike-slip fault at shallow depth (10 km), facilitating strong LAIC through mechanisms such as Radon emanation and atmospheric electric field perturbations.

In contrast, the E2 earthquake (Mw 7.0, October 30, 2020) occurred 15 km offshore in the Aegean Sea with a normal fault mechanism. Oceanic crust in this region has electrical conductivity of ~ 0.01–0.03 S/m, which is 10–100

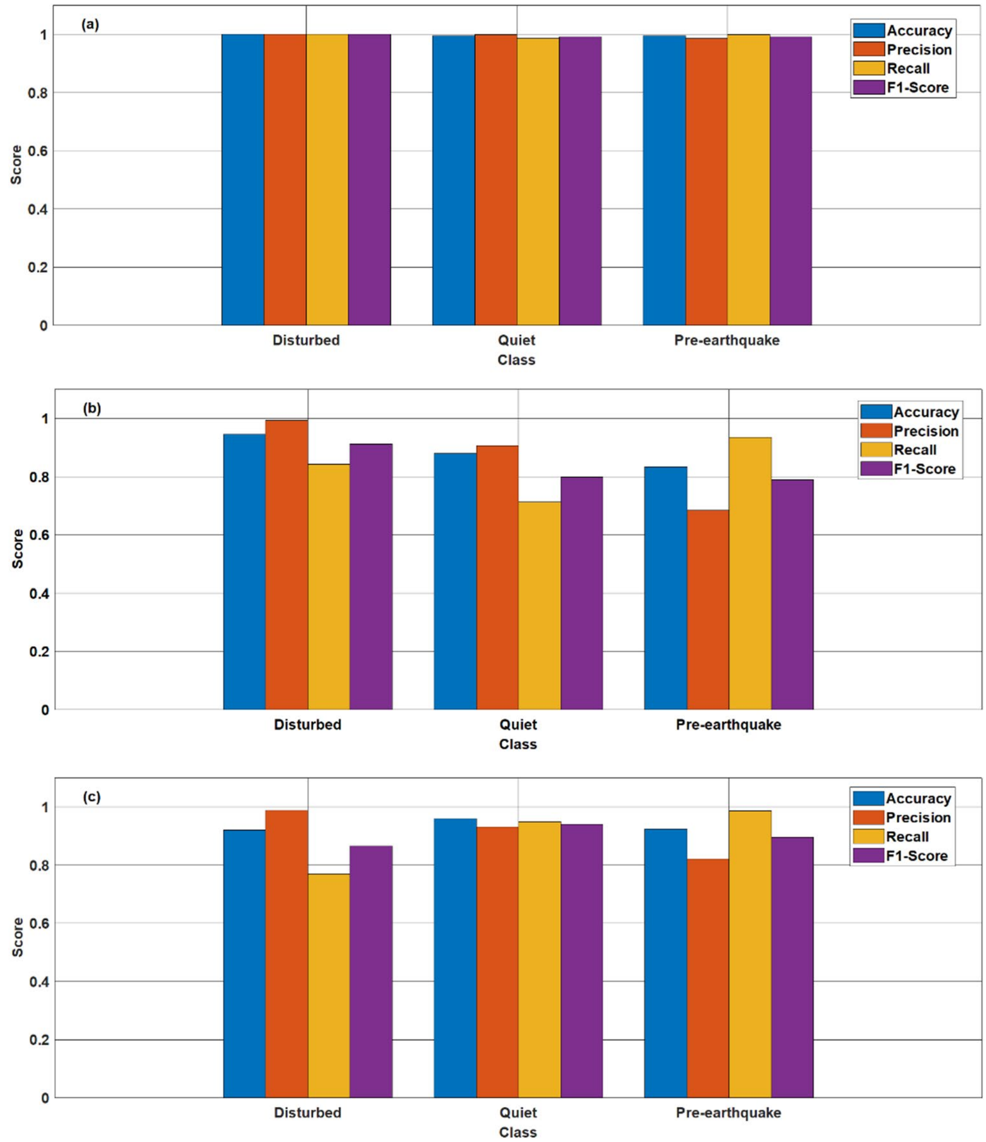


Fig. 9 Performance metrics comparison by class for: (**a**) E1 earthquake (Accuracy: 99.48%, κ = 0.992), (**b**) E2 earthquake (Accuracy: 82.94%, κ = 0.744) and (**c**) E3-E4 earthquakes (Accuracy: 90.21%, κ = 0.853). Detailed values in Supplementary Table S1

times lower than continental crust (~0.1–1.0 S/m) (Filloux 1982; Simpson and Bahr 2005). This low conductivity significantly reduces the strength and penetration depth of seismogenic electric fields into the atmosphere and ionosphere, resulting in weaker and more diffuse LAIC signals compared to continental strike-slip events (E1, E3–E4). Additionally, marine boundary layer effects including high humidity (>75%), sea spray aerosols and enhanced evaporation increase atmospheric conductivity by 2–3 times (Bennett and Harrison 2007) and elevate TEC variability (standard deviation ~ 3.7 TECU in E2 vs. 2.8 TECU in E1 during quiet periods; Table 3). These factors introduce high-frequency noise that masks pre-seismic ionospheric anomalies. Furthermore, the earthquake occurred during moderate geomagnetic activity (Kp = 3–4 in the preceding week), producing ionospheric disturbances that spectrally overlap with seismic precursors. This spectral overlap represents a critical instance of mixed-source signal contamination, where the classifier struggles to disentangle the weak pre-seismic electric field effects inherent to offshore events from moderate geomagnetic forcing. The reduced Precision observed in the İzmir case highlights a central problem in seismo-ionospheric research: even minor space weather events can effectively mask precursor signatures when lithosphere-ionosphere coupling is inefficient, creating an ambiguity that purely statistical classifiers find difficult to resolve without additional physics-based filtering constraints. Together, these effects explain the lower classification Accuracy (82.94%) and increased overlap between Class-0 and Class-2 in the feature space (Fig. 7b).

The E3–E4 earthquake doublet (Mw 7.8 and Mw 7.5, February 6, 2023) represents an intermediate case due to the 9-hour interval between the two events. The first shock (E3, 04:17 local time) occurred within the preparation

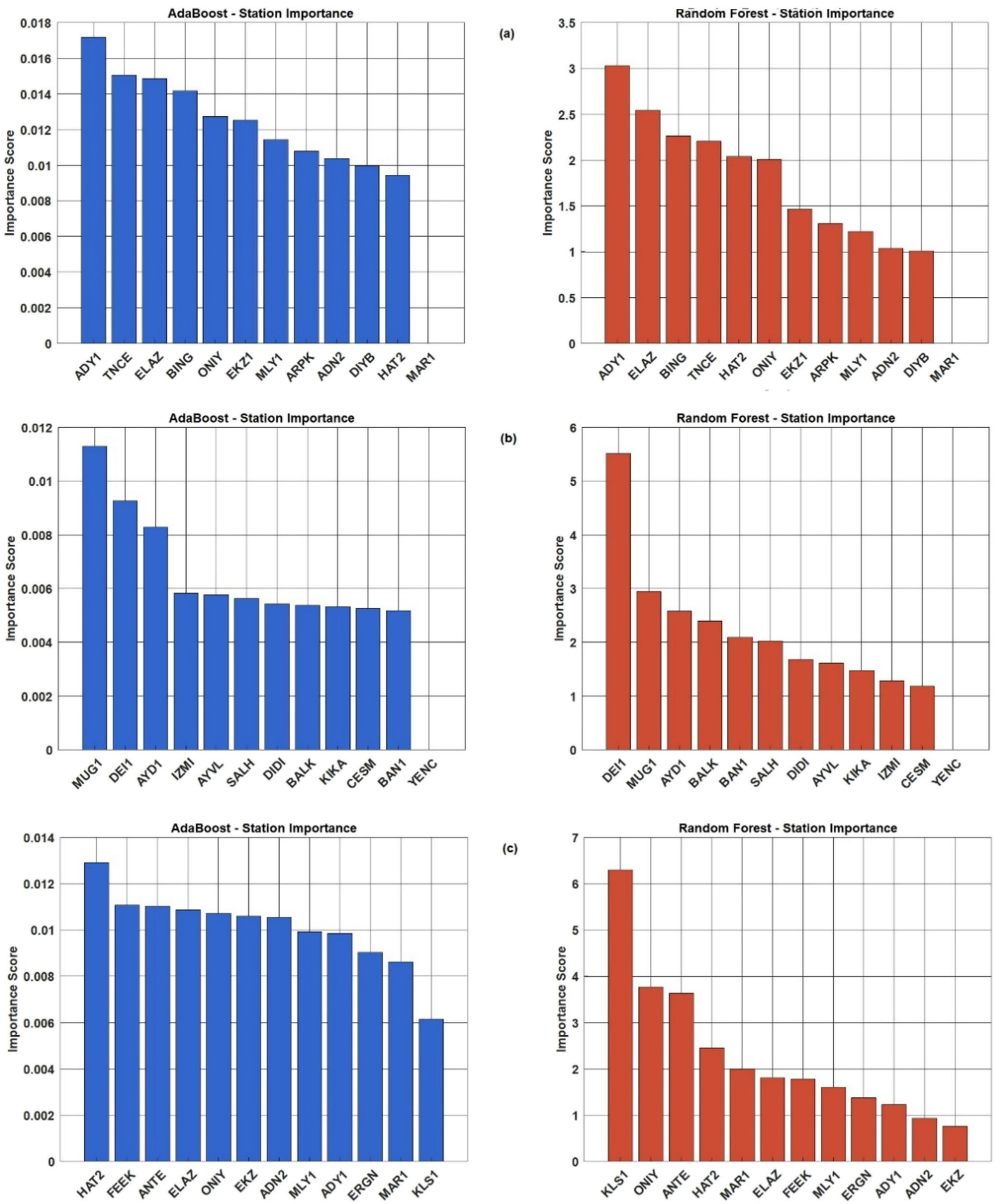


Fig. 10 Relative importance scores of GNSS stations for (**a**) E1, (**b**) E2 and (**c**) E3-E4. Left panels: AdaBoost importance; right panels: RF importance. AdaBoost shows weak-to-moderate inverse correlation with epicentral distance (non-significant), while RF demonstrates moderate inverse correlation (E3-E4: ρ = −0.63, $p = 0.04$)

Table 6 10-fold stratified cross-validation results (mean ± standard deviation). Statistical significance assessed via paired t-test across 10 folds: $p<0.001$, ns = not significant. RF achieves statistically significant improvements for E2 and E3-E4, with performance differences for E1 not reaching significance due to near-perfect Accuracy of both methods

Earthquake	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score	Cohen's κ	*p*-value (paired t-test)
E1 (Elazığ)	AdaBoost	99.49 ± 0.244	99.50 ± 0.239	99.49 ± 0.244	0.99 ± 0.002	0.992 ± 0.003	0.000111503
	Random Forest	100	100	100	1	1	0.000111503
E2 (İzmir)	AdaBoost	83.47 ± 2.090	86.40 ± 1.566	83.47 ± 2.089	0.83 ± 0.002	0.992 ± 0.020	1.223E-09
	Random Forest	99.78 ± 0.380	99.79 ± 0.372	99.78 ± 0.380	0.83 ± 0.380	0.992 ± 0.003	1.223E-09
E3-E4 (Kahramanmaraş)	AdaBoost	88.33 ± 1.534	90.04 ± 1.122	88.33 ± 1.522	0.99 ± 0.002	0.882 ± 0.015	2.471E-09
	Random Forest	99.69 ± 0.330	99.69 ± 0.325	99.69 ± 0.331	0.99 ± 0.003	0.995 ± 0.004	2.471E-09

phase of the second (E4, 13:24 local time), meaning that TEC anomalies observed in the 3-day pre-E3 window likely include cumulative stress buildup from both impending ruptures (Barbot et al. 2023; Melgar et al. 2023). This foreshock–mainshock interaction may enhance ionospheric precursors through superimposed LAIC effects, as stress transfer along the ~300 km rupture zone (East Anatolian Fault) amplifies electric field generation and radon emanation over a broader area. Consequently, pre-E3 TEC signals are not purely representative of a single event but reflect a composite anomaly, which explains the high Class-2 Recall (98.64%) despite moderate overall Accuracy (90.21%). The spatial distribution of station importance (Fig. 10c) further supports this, with peak contributions from stations near the E3 epicenter (MAR1) and E4 nucleation zone (ANTE), indicating dual-source ionospheric excitation. While this complicates isolated precursor detection, it demonstrates that doublet sequences can produce more detectable ionospheric signatures than isolated events of similar magnitude.

Conclusion

This study demonstrates the effectiveness of Adaptive Boosting (AdaBoost) ensemble learning for identifying pre-seismic ionospheric anomalies using IONOLAB-TEC data from Türkiye's TNPGN-Active GNSS network. Three major earthquakes were analyzed: Elazığ (Mw 6.7, 2020), İzmir (Mw 7.0, 2020) and the Kahramanmaraş doublet (Mw 7.8 and 7.5, 2023), using a three-class classification scheme distinguishing geomagnetically quiet (Class-0), disturbed (Class-1) and pre-earthquake (Class-2) conditions.

Based on cross-validated results, this case-study validation suggests RF as the preferred algorithm for seismo-ionospheric classification, achieving 99.69–100.00% Accuracy with lower variance (σ = 0.00–0.38% vs. AdaBoost's 0.24–2.09%) and perfect classification capability for strong continental signals. Although AdaBoost offers faster training (0.69–0.74 s vs. 1.57–1.79 s per event, approximately 50–60% faster), making it a viable alternative for near-real-time applications, RF's superior generalization stability across diverse tectonic settings favors its use in future operational monitoring systems pending multi-event validation. The model achieves an average classification Accuracy of 90.77% with Cohen's kappa values ranging from 0.744 to 0.992, indicating substantial to almost perfect agreement. Statistical validation confirms that performance significantly exceeds baseline classifiers (McNemar's test $p<0.001$, chi-square $p<0.001$). Classification Accuracy varies systematically with earthquake characteristics: the E1 continental strike-slip event achieves 99.48% Accuracy with near-perfect pre-earthquake class metrics, while the E2 offshore normal-fault event yields 82.94% Accuracy with reduced Precision (86.11%) and the E3-E4 continental doublet demonstrates 90.21% Accuracy with strong pre-earthquake performance (F1 = 0.900).

Cross-validated performance comparison (Table 6) establishes RF as the superior algorithm for seismo-ionospheric classification, achieving 99.69–100.00% Accuracy across diverse tectonic settings (continental strike-slip, offshore normal-fault, continental doublet) with minimal variance (σ = 0.00–0.38%). While AdaBoost demonstrates strong performance (83.47–99.49%), representing the first successful application of boosting algorithms to three-class seismo-ionospheric classification with geomagnetic filtering, RF's parallel ensemble strategy provides superior generalization on limited samples ($N=108$ per earthquake). This work establishes critical methodological guidance for the seismo-ionospheric community: k-fold cross-validation must be employed as standard practice for ML-based studies with limited earthquake samples ($N<200$), as proper validation protocols including fold-wise normalization and k-fold cross-validation are essential to ensure reliable performance assessment on limited geophysical samples.

We emphasize that the present study constitutes a case-study validation across three carefully selected earthquake sequences rather than a comprehensive operational assessment. While the consistent RF superiority across contrasting tectonic settings (continental strike-slip, offshore normal-fault, continental doublet) is encouraging, the limited sample of three events precludes definitive claims about global applicability. The path from proof-of-concept to operational

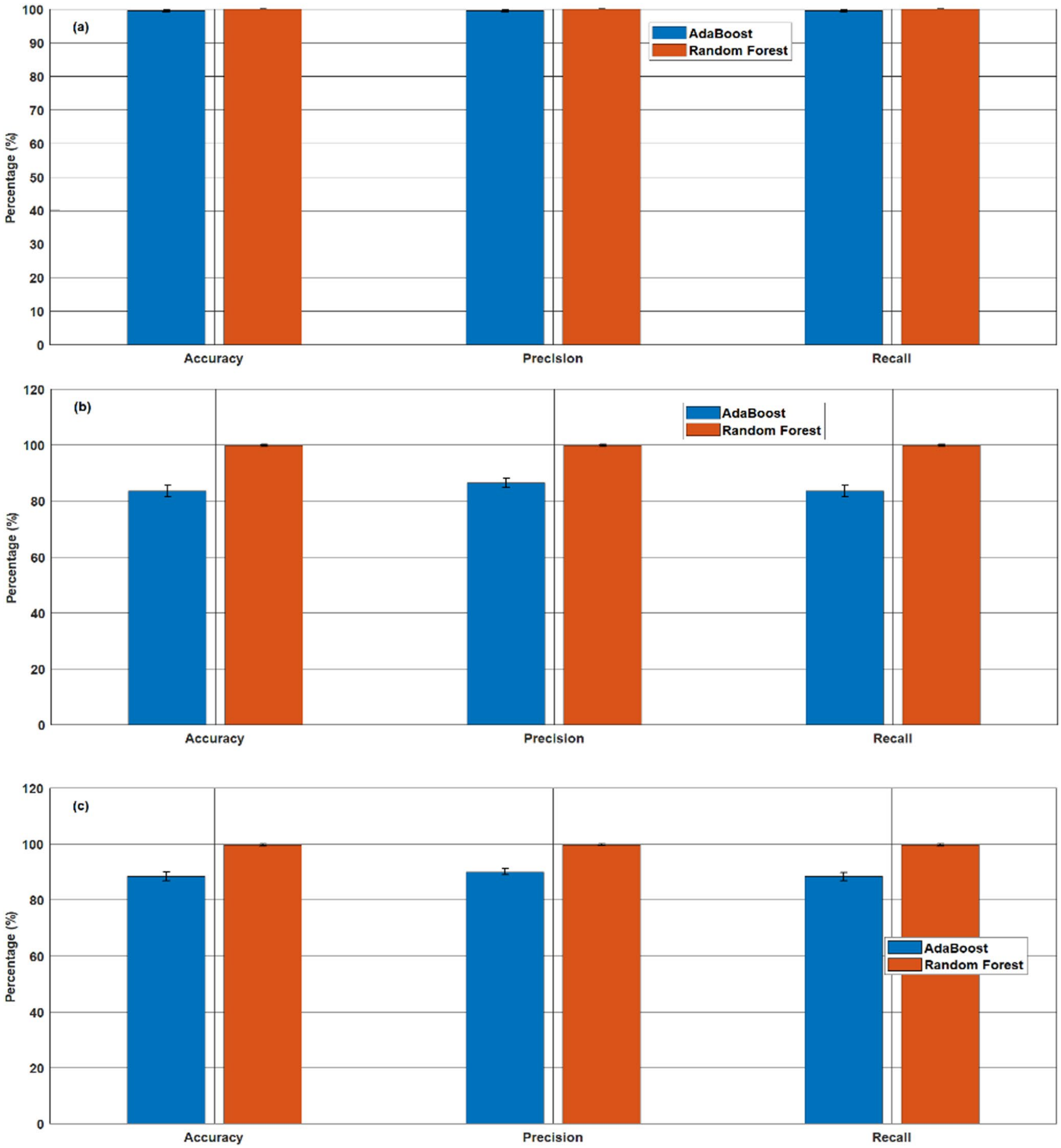


Fig. 11 10-fold stratified cross-validation performance comparison between AdaBoost and RF for: (**a**) E1, (**b**) E2 and (**c**) E3-E4. Bar heights represent mean values across 10 folds; error bars indicate ± 1 standard deviation. RF demonstrates consistently higher performance with lower variance across all tectonic settings

monitoring requires systematic expansion along several axes: (1) multi-event validation incorporating 30–50 earthquakes across the full magnitude spectrum, (2) multi-cycle testing to assess model stability across different solar cycle phases and seasonal ionospheric variability, (3) prospective (forward-looking) evaluation where models trained on historical events are tested on subsequent earthquakes without retraining, and (4) integration with complementary precursor parameters (Radon emissions, thermal anomalies, ULF magnetic signals) within multi-parameter frameworks.

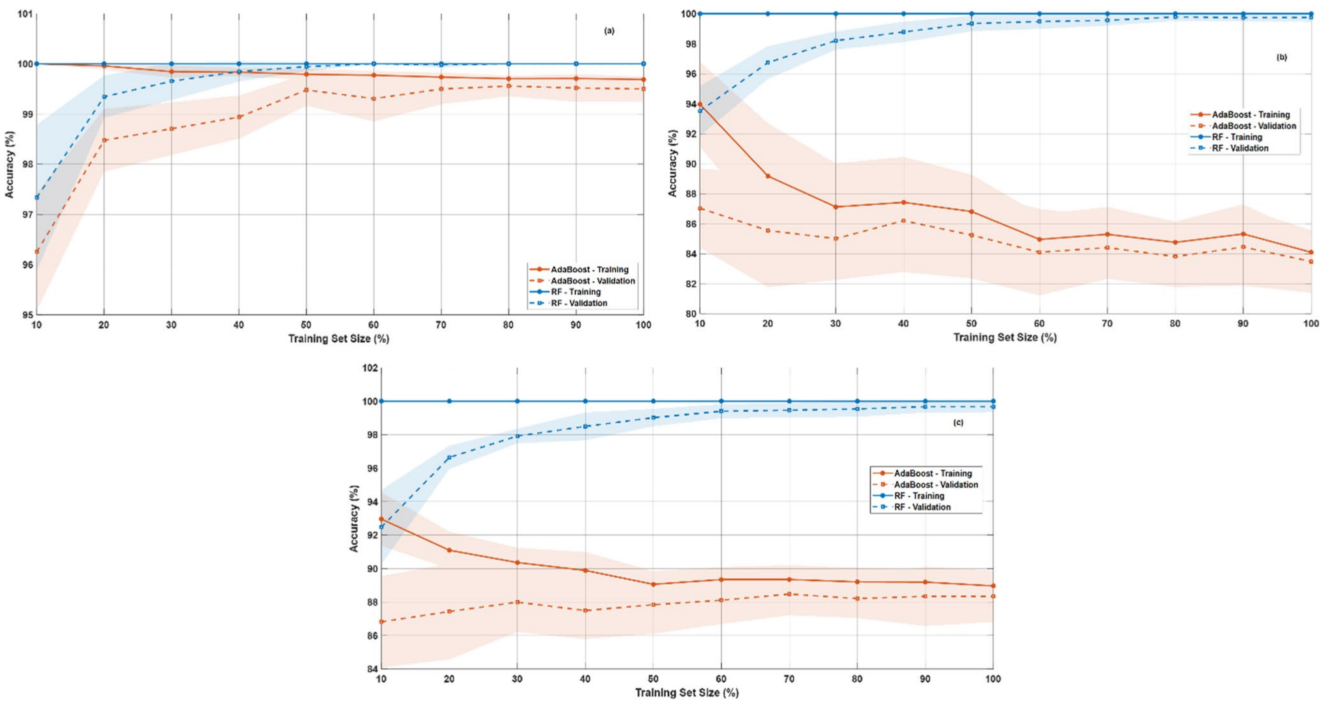


Fig. 12 Learning curves under 10-fold cross-validation for AdaBoost (orange) and RF (blue) across: (**a**) E1, (**b**) E2 and (**c**) E3-E4. Solid lines represent training Accuracy; dashed lines represent validation Accuracy. Shaded regions indicate ± 1 standard deviation across folds

Table 7 Statistical validation metrics for the AdaBoost classification model

Statistical Test	E1 Earthquake	E2 Earthquake	E3-E4 Earthquakes	Overall	Interpretation
Chi-square test (χ^2)	χ^2 = 3060.5, df=4, p<0.001	χ^2 = 1866.7, df=4, p<0.001	χ^2 = 2307.8, df=4, p<0.001	χ^2 = 7235.0, df=12, p<0.001	Classification significantly better than chance
McNemar's test	χ^2 = 1026.0 p<0.001	χ^2 = 634.1 p<0.001	χ^2 = 694.9 p<0.001	χ^2 = 2355.0 p<0.001	Significant improvement over baseline
Cohen's kappa (κ) AdaBoost/RF	κ=0.992/1	κ=0.744/0.99	κ=0.853/0.99	κ=0.863/0.99	Strong (AdaBoost) to perfect (RF) agreement
F1-score (macro-avg) AdaBoost/RF	0.994/1	0.832/0.99	0.900/0.99	0.909/0.993	High (AdaBoost) to near-perfect (RF) quality
Matthews Correlation Coefficient (MCC)	0.992	0.757	0.859	0.869	Strong correlation predicted-actual
Spearman correlation (ρ) *Station importance vs. epicentral distance* AdaBoost/RF	ρ = −0.55, p=0.06/ ρ = −0.37, p=0.22	ρ = −0.31, p=0.31/ ρ = −0.69, p=0.16	ρ=0.10, p=0.75/ ρ = −0.63, p=0.04	ρ = −0.32, p=0.75/ ρ = −0.56, p=0.22	Weak-to-moderate inverse correlation (AdaBoost, non-significant); moderate inverse correlation (RF, partially significant)

Performance variations reflect fundamental differences in seismo-ionospheric coupling efficiency across tectonic settings. Continental strike-slip environments (E1, E3-E4) produce strong, spatially coherent ionospheric signatures enabling exceptional classification, while offshore normal-fault events (E2) generate weaker signals degraded by low oceanic crust conductivity (approximately 0.01–0.05 S/m compared to continental values) and maritime atmospheric conditions. The three-class classification framework successfully distinguishes earthquake-related ionospheric disturbances from space weather effects, reducing false positive rates to 9.11% compared to 12–15% reported in binary classification approaches. The superior Recall for pre-earthquake classes (98.64% average for E3-E4) prioritizes

Table 8 Comparison of ML approaches for TEC-based earthquake anomaly detection

Study	Method	Region	No. of Earthquakes	Magnitude Range	Classes	Accuracy (%)	Precision (%)	Recall (%)	F1-score	False Positive Rate (%)
Akyol et al. (2020)	SVM (EQ-PD)	Italy	Case study	M≥5.0	2 (binary)	~85	~83	~86	0.84	~15
Natras et al. (2022)	RF/AdaBoost/XGBoost Ensemble	Global (multi-latitude)	VTEC forecasting	N/A	Regression	RMSE: 1.09–2.15 TECU	N/A	N/A	R^2: 0.95–0.98	N/A
Jeong et al. (2024)	LSTM & ConvLSTM	Korean Peninsula	Regional TEC forecast	N/A	Regression	RMSE: 1.28–5.28 TECU	N/A	N/A	R^2: 0.94–0.95	N/A
Lin and Chiou (2020)	CNN+Butterworth Filter	Global TEC maps	Multiple events	Various Mw	2 (binary)	~88	~85	~89	0.87	~12
Saqib et al. (2023)	LSTM	Pakistan (Awaran)	1	Mw 7.7	Binary	High (0.07 TECU loss)	N/A	N/A	N/A	N/A
This study (E1)	AdaBoost	Türkiye	1	Mw 6.7	3 (multi-class)	99.48	99.48	99.48	0.99	0.514
This study (E2)	AdaBoost	Türkiye	1	Mw 7.0	3 (multi-class)	82.94	86.11	82.94	0.83	17.05
This study (E3-E4)	AdaBoost	Türkiye	2	Mw 7.5–7.8	3 (multi-class)	90.21	91.36	90.21	0.90	9.78
This study (Average)	AdaBoost	Türkiye	3 events	Mw 6.7–7.8	3 (multi-class)	90.77	92.31	90.77	0.90	9.11

detection sensitivity appropriate for public safety applications where minimizing false negatives is paramount.

Several important limitations constrain the generalizability of these findings and should be addressed in future research. The analysis encompasses only three earthquake sequences within Türkiye during 2020–2023, corresponding to solar minimum conditions. Validation across diverse seismotectonic settings, different fault types, earthquake magnitudes ranging from Mw 5.0 to 8.0+, varied crustal compositions and different phases of the solar activity cycle is essential before global deployment. The 3-day pre-earthquake detection window is selected based on literature precedent showing significant ionospheric anomalies within this timeframe (Liu et al. 2004a, b; Heki et al. 2024). However, this represents a fixed methodological choice that may not be optimal for all earthquake types, magnitudes, or tectonic settings. We acknowledge that no sensitivity analysis comparing alternative window lengths (1-day, 3-day, 5-day) was performed in this study. With only three earthquake events (N=3), such a comparison would require reprocessing all raw IONOLAB-TEC data, redefining class boundaries and rerunning all classification experiments for each window length, a substantial computational undertaking that falls outside the scope of this initial case study and one where the small event count would render statistical conclusions unreliable. Classification performance may be sensitive to this parameter: shorter windows (1-day) risk missing gradually developing precursors that emerge over multiple days, while longer windows (5-day or 7-day) introduce increasing contamination from unrelated ionospheric variability unconnected to seismic activity. We emphasize this as an important methodological limitation. Future work with expanded earthquake catalogs (30–50 events) should systematically compare window lengths through cross-validation to identify optimal detection windows for different tectonic settings and magnitude ranges, and explore adaptive window selection strategies based on earthquake magnitude or fault geometry.

The limited dataset size (108 samples per earthquake) restricts robust hyperparameter optimization through nested cross-validation. While we employ 10-fold CV for performance evaluation, formal hyperparameter tuning via nested CV (outer 10-fold × inner 5-fold) would require 50 model fits per hyperparameter combination, creating substantial risk of overfitting to the specific dataset structure. Our hyperparameters are based on learning curve convergence (Fig. 4), literature precedent (Freund and Schapire 1997; Breiman et al. 2017) and preliminary sensitivity analysis showing < 2–3% performance variation across reasonable ranges. Expansion to 30–50 earthquakes spanning diverse magnitudes (Mw 5.0–8.0+), fault types (thrust, normal, oblique-slip) and solar activity phases would enable exhaustive grid search to establish generalizable optimal configurations and validate whether the RF > AdaBoost pattern observed here holds universally or varies with earthquake characteristics.

Additionally, the clustered structure of the dataset, where multiple stations observe shared ionospheric conditions on the same calendar day, introduces dependencies that standard k-fold cross-validation does not fully account for. While our stratified 10-fold CV provides reliable comparative assessment between algorithms, absolute performance estimates may be optimistic relative to leave-one-day-out or leave-one-station-out validation designs that would more rigorously test temporal and spatial generalization, respectively. Future studies with expanded earthquake catalogs should employ grouped cross-validation strategies to obtain more conservative and operationally realistic performance estimates.

While Class-1 reference days are selected to capture strong geomagnetic storms, residual space weather activity during pre-earthquake periods (Kp values of 2–3) introduces classification challenges, highlighting that perfect geomagnetic filtering remains elusive in operational scenarios. Additionally, the use of a single reference period for geomagnetic disturbances (April 2023) limits the assessment of the model's generalization across weaker or different types of space weather events. Similarly, the use of a single severe geomagnetic storm as the common disturbed-day reference provides a controlled comparison framework but does not represent the full spectrum of geomagnetic noise conditions. Operational deployment would require training on multiple storm types spanning a range of intensities and morphologies. Current station selection criteria combine proximity, data quality and EPZ considerations but lack systematic optimization, with quantitative selection thresholds requiring validation across diverse network geometries and earthquake locations. Furthermore, the reduced classification Precision observed for the İzmir event underscores the limitation of the current framework in handling mixed-source signals; concurrent minor geomagnetic activity can contaminate the pre-earthquake window, leading to false positives that remain a persistent challenge for operational forecasting in offshore tectonic settings.

In conclusion, within the constraints of this preliminary case-study analysis, AdaBoost ensemble learning is established as a promising tool for automated detection of pre-seismic ionospheric anomalies in IONOLAB-TEC data. Future research directions should include expansion of the training corpus to 30–50 earthquakes across global tectonic settings with systematic sampling of magnitude, depth and fault-type variations. Controlled synthetic experiments varying station network geometry, noise levels and earthquake source parameters would isolate algorithm-specific versus data-driven performance factors. Development of physics-based simulation frameworks for testing classification robustness under known signal conditions would complement empirical studies. Integration of additional geophysical parameters including atmospheric electric fields, Radon emanation and thermal radiation in multi-parameter ensemble frameworks represents a promising avenue for improving detection reliability. Systematic hyperparameter optimization through nested cross-validation with expanded datasets, validation of proposed station selection metrics across diverse network configurations and exploration of temporal dynamics beyond 3-day windows through adaptive window selection based on earthquake magnitude forecasts constitute important next steps toward operational earthquake precursor monitoring systems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12145-026-02109-7.

Acknowledgements This research is supported by TUBİTAK EEEAG Grant 124E027 and was carried out within the scope of Kastamonu University Electrical and Electronics Engineering Department EEM401 Electrical and Electronics Engineering Design course. The authors are grateful to anonymous reviewers for their comments and contributions, which have been very helpful and constructive for the authors in improving the paper. The authors also wish to thank Prof. Dr. Feza Arikan and IONOLAB group for their outstanding efforts on IONOLAB-BIAS and IONOLAB-TEC Algorithm.The IONOLAB-TEC data used in this study were obtained from the IONOLAB Group (www.ionolab.org) and are available upon reasonable request to the corresponding author. The GIM-TEC, Satellite DCB and ephemeris data that are used in computation of IONOLAB-TEC are obtained from IGS Analysis Center of Jet Propulsion Laboratory (JPL) at ftp://cddis.gsfc.nasa.gov/pub/gps/products/ionex. TNPGN-Active RINEX data set can be obtained by the permission from General Directorate of Land Registry and Cadastre, Department of Mapping at https://www.tusaga-aktif.gov.tr/.

Author contributions Secil Karatay: Conceptualization, Methodology, Software (algorithm code development), Formal analysis, Data curation, Writing—original draft, Writing—review & editing, Visualization, Supervision, Project administration. Zeynep Mantaroglu: Methodology (documentation), Writing—original draft, Data processing. Esra Nur Serbetli: Methodology, Writing—original draft, Data processing. All authors have read and approved the final manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). It's not funded by any agency/organization either technically or financially.

Data availability The classification algorithms and processing scripts used in this research can be shared with interested researchers upon request to the corresponding author for validation and replication purposes. Implementation follows standard ensemble learning frameworks described in Freund & Schapire (1997) and Breiman et al. (2017). Additionally, any derived data supporting the findings of this study can be provided upon request.

Declarations

Ethical Approval All authors have seen and agreed with the contents of the manuscript and are looking forward to publishing this paper in this journal.

Consent for publication All authors gave explicit consent to publish this manuscript.

Competing interests The authors declare no competing interests.

References

Aktug B, Ozener H, Dogru A et al (2016) Slip rates and seismic potential on the East Anatolian Fault System using an improved GPS velocity field. J Geodyn 94–95:1–12. https://doi.org/10.1016/j.jog.2016.01.001

Akyol A, Arikan O, Arikan F (2020) A machine learning-based detection of earthquake precursors using ionospheric data. Radio Sci 55(11):1–21. https://doi.org/10.1029/2019RS006931

Arikan F, Erol CB, Arikan O (2003) Regularized estimation of vertical total electron content from Global Positioning System data. J Geophys Res Space Phys 108(A12):1–20. https://doi.org/10.1029/2002JA009605

Arikan F, Erol CB, Arikan O (2004) Regularized estimation of VTEC from GPS data for a desired time period. Radio Sci 39(6):1–10. https://doi.org/10.1029/2004RS003061

Asaly S, Gottlieb L-A, Inbar N, Reuveni Y (2022) Using support vector machine (SVM) with GPS ionospheric TEC estimations to potentially predict earthquake events. Remote Sens 14(12):1–17. https://doi.org/10.3390/rs14122822

Asim KM, Idris A, Iqbal T, Martínez-Álvarez F (2018) Seismic indicators based earthquake predictor system using genetic programming and AdaBoost classification. Soil Dyn Earthq Eng 111:1–7. https://doi.org/10.1016/j.soildyn.2018.04.020

Astafyeva E (2019) Ionospheric detection of natural hazards. Rev Geophys 57(4):1265–1288. https://doi.org/10.1029/2019RG000668

Barbot S, Luo H, Wang T, Hamiel Y, Piatibratova O, Javed MT, Braitenberg C, Gurbuz G (2023) Slip distribution of the February 6, 2023 Mw 7.8 and Mw 7.6, Kahramanmaraş, Turkey earthquake sequence in the East Anatolian Fault Zone. Seismica 2(3):1–17. https://doi.org/10.26443/seismica.v2i3.502

Bennett AJ, Harrison RG (2007) Atmospheric electricity in different weather conditions. Weather 62(10):277–283. https://doi.org/10.1002/wea.97

Berhich A, Belouadha F-Z, Kabbaj MI (2022) A location-dependent earthquake prediction using recurrent neural network algorithms. Soil Dyn Earthq Eng 161:1–25. https://doi.org/10.1016/j.soildyn.2022.107389

Breiman L, Friedman J, Olshen R (2017) Classification and Regression Trees. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9781315139470

Budak C, Karatay S, Erken F, Cinar A (2024) Classification of the ionospheric disturbances caused by geomagnetic and seismic activity with K-nearest neighbors algorithm. Wireless Pers Commun 134:1551–1569. https://doi.org/10.1007/s11277-024-10965-z

Convertito V, Giampaolo F, Amoroso O, Piccialli F (2024) Deep learning forecasting of large induced earthquakes via precursory signals. Sci Rep. https://doi.org/10.1038/s41598-024-52935-2

De Santis A, Marchetti D, Pavón-Carrasco FJ, Cianchini G, Perrone L et al (2019) Precursory worldwide signatures of earthquake occurrences on Swarm satellite data. Sci Rep 9:1–13. https://doi.org/10.1038/s41598-019-56599-1

Dietterich TG (1998) Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 10(7):1895–1923. https://doi.org/10.1162/089976698300017197

Dobrovolsky IP, Zubkov SI, Miachkin VI (1979) Estimation of the size of earthquake preparation zones. Pure Appl Geophys 117:1025–1044. https://doi.org/10.1007/BF00876083

Domingos P (2012) A few useful things to know about machine learning. Commun ACM 55(10):78–87. https://doi.org/10.1145/2347736.2347755

Drucker H (1997) Improving regressors using boosting techniques. *4th International Conference on Machine Learning (ICML '97)* (pp. 107–115). Nashville: Morgan Kaufmann Publishers

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27(8):861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Filloux JH (1982) Magnetotelluric experiment over the ROSE area. J Geophys Res Solid Earth 87(B10):8364–8378. https://doi.org/10.1029/JB087iB10p08364

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139. https://doi.org/10.1006/jcss.1997.1504

GFZ (1956) *GFZ Helmholtz Centre for Geosciences*. Retrieved 10 27, 2025, from https://www-app3.gfz-potsdam.de/kp_index/Kp_ap_since_1932.txt

Haider S, Shah M, Li B, Jamjareegulgarn P, de Oliveira-Júnior J, Zhou C (2025) Synchronized and co-located ionospheric and atmospheric anomalies associated with the 2023 Mw 7.8 Turkey earthquake. Remote Sens 16(2):1–18. https://doi.org/10.3390/rs16020222

Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second ed. Springer, New York

Heki K, Nakatani M, Zhan. W (2024) Ionospheric changes immediately before the 2008 Wenchuan earthquake. Adv Space Res 73(9):4539–4545. https://doi.org/10.1016/j.asr.2024.02.011

IONOLAB (2006) *Ionosphere Research Laboratory*. Retrieved 10 13, 2025, from https://www.ionolab.org/

IZMIRAN (1939) *Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation Russian Academy of Sciences*. Retrieved 12 13, 2025, from https://www.izmiran.ru/

Jain A, Duin R, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37. https://doi.org/10.1109/34.824819

Jeong S-H, Lee WK, Kil H, Jang S, Kim J-H, Kwak Y-S (2024) Deep learning-based regional ionospheric total electron content prediction—long short-term memory (LSTM) and convolutional LSTM approach. Space Weather 22(1):1–10. https://doi.org/10.1029/2023SW003763

Joshi S, Kannaujiya S, Joshi U (2023) Analysis of GNSS data for earthquake precursor studies using IONOLAB-TEC in the Himalayan region. Quaternary 6(2):1–13. https://doi.org/10.3390/quat6020027

Karatay S, Arikan F, Arikan O (2010) Investigation of total electron content variability due to seismic and geomagnetic disturbances in the ionosphere. Radio Sci 45(5):1–12. https://doi.org/10.1029/2009RS004313

Karatay S, Arikan F, Pirti A (2025) Spatio-temporal evaluation of ionospheric disturbances before, during and after earthquakes using differential rate of TEC (DROT) from GPS measurements. Environ Earth Sci 84(1–15). https://doi.org/10.1007/s12665-025-12216-1

Karatay S, Gul SE (2023) Prediction of GPS-TEC on Mw > 5 earthquake days using Bayesian regularization backpropagation algorithm. IEEE Geosci Remote Sens Lett 20:1–5. https://doi.org/10.1109/LGRS.2023.3262028

Kaushal A, Gupta AK, Sehgal VK (2025) Earthquake prediction optimization using deep learning hybrid RNN-LSTM model for seismicity analysis. Soil Dyn Earthq Eng 195:1–23. https://doi.org/10.1016/j.soildyn.2025.109432

Kiruthiga S, Mythili S (2024) Prediction of ionospheric TEC during the occurrence of earthquakes in Indonesia using ARMA and CoK models. Front Astron Space Sci 11:1–15. https://doi.org/10.3389/fspas.2024.1415323

Kohavi R (1995) Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. In A. s.-v. model (Ed.), *14th International Joint Conference on Artificial Intelligence*, *2*, pp. 1137–1143

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174. https://doi.org/10.2307/2529310

Li G, Jiang H, Liu J, Liu C, Zhang Y (2024) An emergency observation method for space debris dangerous conjunction events based on nanosatellite constellation. Adv Space Res 73(11):5545–5559. https://doi.org/10.1016/j.asr.2023.07.052

Lin J-W (2022) Predicting ionospheric precursors before large earthquakes using neural network computing and the potential development of an earthquake early warning system. Nat Hazards 113:1519–1542 . https://doi.org/10.1007/s11069-022-05356-1

Lin J-W, Chiou J-S (2020) Detecting total electron content precursors before earthquakes by examining total electron content images based on butterworth filter in convolutional neural networks. IEEE Access 8:110478–110494. https://doi.org/10.1109/ACCESS.2020.3001337

Liu C-Y, Liu JY, Chen Y-I, Qin F (2018) Statistical analyses on the ionospheric total electron content related to M≥6.0 earthquakes in China during 1998–2015. Terr Atmospheric Ocean Sci 29(5):485–498. https://doi.org/10.3319/TAO.2018.03.11.01

Liu JY, Chen YI, Chuo YJ, Tsai HF (2001) Variations of ionospheric total electron content during the Chi-Chi earthquake. Geophys Res Lett 28(7):1383–1386. https://doi.org/10.1029/2000GL012511

Liu JY, Chen YI, Jhuang HK, Lin YH (2004a) Ionospheric foF2 and TEC anomalous days associated with M>=5.0 earthquakes in Taiwan during 1997–1999. Terr Atmos Ocean Sci 15(3):371–383. https://doi.org/10.3319/TAO.2004a.15.3.371(EP)

Liu JY, Chuo YJ, Shan SJ, Tsai YB, Chen YI, Pulinets SA, Yu SB (2004) Pre-earthquake ionospheric anomalies registered by continuous GPS TEC measurements. Ann Geophys 22(5):1585–1593. https://doi.org/10.5194/angeo-22-1585-2004

Maletckii B, Astafyeva E, Sanchez SA, Kherani EA, Paula ER (2023) The 6 February 2023 Türkiye earthquake sequence as detected in the ionosphere. Journal of Geophysical Research: Space Physics 128(9):1–11. https://doi.org/10.1029/2023JA031663

Marrissey LS, Rahnamoun A, Nakhla S (2020) The effect of atomic oxygen flux and impact energy on the damage of spacecraft metals. Adv Space Res 66(6):1495–1506. https://doi.org/10.1016/j.asr.2020.05.021

Melgar D, Taymaz T, Ganas A, Crowell BW, Öcalan T (2023) Sub- and super-shear ruptures during the 2023 Mw 7.8 and Mw 7.6 earthquake doublet in SE Türkiye. Seismica 2(3):1–10. https://doi.org/10.26443/seismica.v2i3.387

Mukesh R, Dass SC, Vijay M et al (2024) Analysis of TEC variations and prediction of TEC by RNN during Indonesian earthquakes occurred from 2004 to 2024 and comparison with IRI-2020 model. *Advances in Space, 74*(11), 6100–6118. https://doi.org/10.1016/j.asr.2024.07.055

Natras R, Soja B, Schmidt M (2022) Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. Remote Sens 14(15):1–34. https://doi.org/10.3390/rs14153547

Nayir H, Arikan F, Arikan O, Erol CB (2007) Total Electron Content estimation with Reg-Est. J Geophys Res Space Phys. https://doi.org/10.1029/2007JA012459

Oyama K-I, Devi M, Ryu K, Chen CH, Liu JY et al (2016) Modifications of the ionosphere prior to large earthquakes: report from the Ionosphere Precursor Study Group. Geosci Lett 3:1–10. https://doi.org/10.1186/s40562-016-0038-3

Pirti A, Karatay S (2025) Multi-parameter GNSS analysis reveals atmospheric, ionospheric and crustal precursors seven days before the 2016 Mw 7.8 Ecuador megathrust earthquake. Earth Sci Inf 18:1–22. https://doi.org/10.1007/s12145-025-02037-y

Pulinets SA, Legen'ka AD, Alekseev VA (1994) Pre-Earthquake Ionospheric Effects and their Possible Mechanisms. Dusty and Dirty Plasmas, Noise, and Chaos in Space and in the Laboratory. Springer, Boston, pp 545–557. https://doi.org/10.1007/978-1-4615-1829-7_46

Pulinets S, Herrera VM (2024) Earthquake precursors: the physics, identification, and application. Geosciences 14(8):1–33. https://doi.org/10.3390/geosciences14080209

Pulinets S, Ouzounov D (2011) Lithosphere–atmosphere–ionosphere coupling (LAIC) model – an unified concept for earthquake precursors validation. J Asian Earth Sci 41(4–5):371–382. https://doi.org/10.1016/j.jseaes.2010.03.005

Pulinets S, Ouzounov D, Karelin A, Boyarchuk K (2022) Earthquake Precursors in the Atmosphere and Ionosphere, 1. ed. Springer, Berlin. https://doi.org/10.1007/978-94-024-2172-9

Ridzwan NS, Yusoff SH (2023) Machine learning for earthquake prediction: a review (2017–2021). Earth Sci Inf 16:1133–1149. https://doi.org/10.1007/s12145-023-00991-z

Saied M, Guirguis S, Madbouly M (2023) A comparative study of using boosting-based machine learning algorithms for IoT network intrusion detection. Int J Comput Intell Syst 16:1–15. https://doi.org/10.1007/s44196-023-00355-x

Saqib M, Adil MA, Freeshah M (2023) Pre-earthquake ionospheric perturbation analysis using deep learning techniques. Adv Geomatics 1(1):48–67. https://doi.org/10.5281/zenodo.10202326

Sezen U, Arikan F, Arikan O, Ugurlu O, Sadeghimorad A (2013) Online, automatic, near-real time estimation of GPS-TEC: IONOLAB-TEC. Space Weather 11(5):291–305. https://doi.org/10.1002/swe.20054

Shah M, Aibar AC, Tariq MA, Ahmed J, Ahmed A (2020) Possible ionosphere and atmosphere precursory analysis related to Mw > 6.0 earthquakes in Japan. Remote Sens Environ 239:1–9. https://doi.org/10.1016/j.rse.2019.111620

Shan W, Zhang M, Wang M, Chen H, Zhang R et al (2022) EPM–DCNN: earthquake prediction models using deep convolutional neural networks. Bull Seismol Soc Am 112(6):2933–2945. https://doi.org/10.1785/0120220058

Shen X, Zhang X, Yuan S, Wang L, Cao J et al (2018) The state-of-the-art of the China Seismo-Electromagnetic Satellite mission. Sci China Technol Sci 61:634–642. https://doi.org/10.1007/s11431-018-9242-0

Simpson F, Bahr K (2005) Practical Magnetotellurics. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511614095

Solís-Alulima B, Abati J, López-Carmona A, Gutiérrez-Alonso G, Fernández-Suárez J, Stockli DF (2023) Hf isotopes in detrital zircon grains of the Sierra Albarrana Domain (SW Iberian Massif): eburnean v. archean basement signatures. J Geol Soc 180(3):1–7. https://doi.org/10.1144/jgs2022-151

Srivastava S, Chandran A, Manta F, Taisne B (2021) GNSS tec-based detection and analysis of acoustic-gravity waves from the 2012 Sumatra double earthquake sequence. J Geophys Res Space Phys 126(6):1–14. https://doi.org/10.1029/2020JA028507

Ullah M, Islam F, Alarifi N, Youssef YM, Anees Y et al (2025) Precursory anomalies in the seismo-ionospheric-atmospheric domain associated with offshore earthquakes of magnitude greater than 7.0. Geomatics Nat Hazards Risk 16. https://doi.org/10.1080/19475705.2025.2555740

USGS (1879) *United States Geological Survey*. Retrieved 07 26, 2025, from https://www.usgs.gov/

Uyanık H, Şentürk E, Akpınar M, Ozcelik S, Kokum M, Freeshah M, Sengur A (2023) A multi-input convolutional neural networks model for earthquake precursor detection based on ionospheric total electron content. Remote Sens 15(24):1–19. https://doi.org/10.3390/rs15245690

Vesnin A, Yasyukevich Y, Perevalova N, Şentürk E (2023) Ionospheric response to the 6 February 2023 Turkey–Syria earthquake. Remote Sens 15(9):1–16. https://doi.org/10.3390/rs15092336

WDC (1999) *WDC for Geomagnetism, Kyoto*. Retrieved 10 27, 2025, from https://wdc.kugi.kyoto-u.ac.jp///dstae///wwwtmp/WWW_dstae03340870.dat

Woith H (2015) Radon earthquake precursor: a short review. Eur Phys J Special Top 224:611–627. https://doi.org/10.1140/epjst/e2015-02395-9

Xie T, Chen B, Wu L, Dai W, Kuang C (2021) Detecting seismo-ionospheric anomalies possibly associated with the 2019 Ridgecrest (California) earthquakes by GNSS, CSES, and Swarm observations. JGR Space Phys 126(9):1–27. https://doi.org/10.1029/2020JA028761

Xiong P, Long C, Zhou H, Zhang X, Shen X (2022) GNSS tec-based earthquake ionospheric perturbation detection using a novel deep learning framework. IEEE J Sel Top Appl Earth Observations Remote Sens 15:4248–4263. https://doi.org/10.1109/JSTARS.2022.3175961

Yaşar M, Kurt K, Yeşil A (2025) A new approach on the complex diffusion tensor in the ionospheric F-region with low latitudes. Geomag Aeron. https://doi.org/10.1134/S0016793225600092

Zhang R, Li H, Shen Y, Yang J, Li W, Zhao D, Hu A (2025) Deep learning applications in ionospheric modeling: progress, challenges, and opportunities. Remote Sens 17(1):1–24. https://doi.org/10.3390/rs17010124

Zorkun NG, Oncul MH, Karatay S, Erken F, Budak C (2025) Identification of the Ionospheric Disturbances Prior to Large Earthquakes in Türkiye Using Random Forests Algorithm. Adv Space Res In Press. https://doi.org/10.1016/j.asr.2025.06.080

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Secil Karatay[1] · **Zeynep Mantaroglu**[1] · **Esra Nur Serbetli**[1]

✉ Secil Karatay
skaratay@kastamonu.edu.tr

Zeynep Mantaroglu
zeynepmr0@gmail.com

Esra Nur Serbetli
esranurserbetli@gmail.com

[1] Department of Electrical Electronics Engineering, Kastamonu University, Kastamonu 37150, Türkiye