Wireless Personal Communications (2024) 134:1551–1569
https://doi.org/10.1007/s11277-024-10965-z

Classification of the Ionospheric Disturbances Caused by Geomagnetic and Seismic Activity with K-Nearest Neighbors Algorithm

Cafer Budak[1] · Secil Karatay[2] · Faruk Erken[2] · Ali Cinar[2]

Accepted: 10 March 2024 / Published online: 6 April 2024

Abstract

Detection of earthquake-precursor signals a few days before the earthquake day has become an area of increasing interest. In recent years, it has been observed that the major earthquakes and geomagnetic activity can cause significant disturbances and anomalies in the ionospheric parameters such as Total Electron Content (TEC). TEC provides important information about the detection of anomalies and disturbances related to seismic and geomagnetic activity in the ionosphere. The main goal of this study is to classify the disturbances due to the seismic and geomagnetic activity in the ionosphere using TEC data. For this purpose, the K-Nearest Neighbors (K-NN) algorithm is applied to TEC estimated from Global Positioning System stations during five earthquakes with magnitudes Mw greater than 5.6 between 1999 and 2016 and for the geomagnetically quiet and disturbed conditions of the ionosphere. The data is divided into four classes as non-earthquake and non-geomagnetic activity, geomagnetic activity, possible earthquake precursor and earthquake for each earthquake. The study is performed into two groups as Group I, where five days before the earthquake day are marked as precursor, and Group II, where three days are marked as precursors. It is observed that 5055 test samples out of a total of 5184 are classified as true whereas 129 are classified as false for Group I. 3912 are classified as true and 105 are classified as false for Group II. For the possible earthquake class, the Accuracy values increase inversely with the distance of the stations from the epicenter and directly related to the magnitude of the earthquakes.

✉ Secil Karatay
skaratay@kastamonu.edu.tr

Cafer Budak
cafer.budak@dicle.edu.tr

Faruk Erken
ferken@kastamonu.edu.tr

Ali Cinar
acinar@kastamonu.edu.tr

[1] Department of Electrical Electronics Engineering, Dicle University, 21100 Diyarbakır, Turkey

[2] Department of Electrical Electronics Engineering, Kastamonu University, 37150 Kastamonu, Turkey

Keywords K-nearest neighbors · Disturbance · Classification · Earthquake · Ionosphere

1 Introduction

To reduce the likelihood of earthquake damage situations and potential risks, it is essential to identify, classify, and predict the precursors before the strong earthquakes. The precursor signals during the major earthquakes are detected and classified by the deterministic models [1–4]. In these models, the signals have probabilities in terms of location, time and magnitude of the earthquakes. The studies based on predicting, classifying and modeling the earthquake-precursor signals are very recent in the literature. These studies progressed after the second half of this century have grown in parallel to the multidisciplinary field such as seismology [5], geology [6], physics [7, 8], remote sensing [9] and signal processing [2] and etc. The studies based on the detection, classification and prediction of earthquake-precursor signals have presupposed on the techniques of observing and measuring some variations in the Earth's crust and then its coupling to the atmosphere couple days before the maximum shock before the breakdown of the fault. Significant results have been observed with strategies based on forecasting, categorizing, and modeling earthquake-precursor signals. They are all still in the surveying and development phases as of right now. Most of the studies related to detection, classification and prediction of earthquake-precursor signals are based on the empirical models that aim to detect the precursor signals a few days before the strong earthquakes. However, the behavior of these precursor signals over long time periods is still an area of research and interest. Therefore, there are still serious discussion regarding the reproducibility and reliability of these techniques.

In order to prevent the damaging impacts of earthquakes, it is important to identify, categorize, and predict the precursor signals before strong earthquakes. Before and during earthquakes, the movements of the earth's crust result in not only quakes but also significant variations that human beings cannot sense directly. In recent years, these variations have been modelled using different ionospheric parameters such as the critical frequency of the ionosphere and the Total Electron Content (TEC). A crucial layer of the Earth's atmosphere, the ionosphere, is situated between 60 and 1000 km above the surface. It profoundly affects satellite and communication systems, near-Earth space events, Space Weather, and the magnetic field of the Earth. The ionosphere is a dispersive and inhomogeneous plasma medium that varies multiple scale spatially and temporally [10]. The spatio-temporal variability of the ionosphere incurs significant disturbed and anomalous signals in High Frequency (HF) communication, satellite systems and satellite based navigation systems [11, 12], control and radar system. The ionosphere's spatiotemporal variability, which is one of the characteristics of the atmosphere, is precisely related to solar, geomagnetic, and seismic activities. It is necessary to observe and model spatio-temporal variations in the ionosphere brought on by coupling of solar, geomagnetic, and seismic events for the welfare and continuation of satellite-based navigation systems [13]. One of the most common ionospheric parameters, TEC is estimated from the ionospheric electron density distribution and affected from the spatio-temporal variations caused by solar, geomagnetic, and seismic activities. TEC is described as the sum of the free electrons throughout a ray path of the radio waves between the satellite and the receiver [14, 15]. TEC is equal to 10^{16} electrons per square meter and its unit is TECU.

Within Earth's upper atmosphere, there is a complex and dynamic interplay between geomagnetic activity and ionospheric disturbances. Geomagnetic activity caused by fluctuations in the solar wind and the Earth's magnetic field significantly influences the ionosphere. Ionospheric disturbances may result from increased solar wind energy during times of increased geomagnetic activity, such as during geomagnetic storms. These disturbances appear as variations in the electron density, anomalies in the ionospheric layers, and the anomalies in the ionospheric plasma [10, 16]. Understanding space weather phenomena requires an understanding of the relationship between geomagnetic activity and the ionosphere. The importance of tracking and forecasting geomagnetic conditions for a range of technology applications on Earth and in space is highlighted by the impact that ionospheric disturbances can have on satellite communication, radio wave propagation, and navigation systems [11, 12]. When the effects of variables like fluctuating solar irradiance and background seasonal changes are thoroughly eliminated, the response tendencies of low- to high-latitude foF2 to rising geomagnetic activity are observed in [17]. Also, [18] has indicated that the contribution of geomagnetic activity to the perturbation of the ionosphere is highly substantial during the solar minimum and ascending phase of the solar cycle, but it is very small when geomagnetic activity is at its highest point, which often occurs during the declining phase.

In recent years, one of the topics studied intensively in the literature is the Lithosphere-Atmosphere–Ionosphere Coupling (LAIC) model [19, 20] that explains the coupling between seismic activity and the electromagnetic condition of the ionosphere. These investigations, which have a recent history, indicate that there is a significant relation between the strong earthquakes and anomalous signals in the ionosphere. The first studies of ionosphere effects about this coupling have been worked just after the Great Alaskan earthquake on Good Friday, March 27, 1964 [21–26]. The coupling of Acoustic Gravity Waves (AGW) with the ionosphere is the first idea put forward for this coupling mechanism between the ionosphere and seismicity [27, 28]. The first studies [29] and [30] have suggested that the critical frequency of the E region of the ionosphere varies during the strong earthquakes as seismic precursors. Also, Very Low Frequency (VLF) has been accepted as seismic ionospheric precursors in this coupling. [31] has introduced that VLF sounds recorded on satellites before the major earthquakes are not produced in the seismic areas and do not interpenetrate to the ionosphere. In the studies about for the seismo-ionospheric coupling model [7, 8, 32–35], the variability of the electron density of the ionosphere obtained from the ionosonde and the Global Positioning System (GPS) networks has been studied for the ionospheric disturbed and quiet conditions. In these studies, some anomalies on the ionospheric TEC have been observed for the stations within the impact area of the earthquake 1 to 5 days before the earthquake. Recently, in [2], a new a new technique, Support Vector Machine (SVM) classifier, based on the Machine Learning has been applied on GPS-TEC data with geomagnetic indices for the detection of earthquake precursors. During the validation period of 266 days, the SVM classifier detected seismic precursors in 17 out of 21 earthquakes while producing 7 false alarms, and in 22 out of 24 earthquakes while generating 13 false alerts during the test period of 282 days.

The majority of analytical studies to find the ionospheric precursor have used the statistical techniques. In these studies [8, 32–34, 36–41], it has been tried to reveal the statistical relations between the ionospheric parameters such as TEC, critical frequency of the ionospheric F2-region. There is still a great deficiency and demand in the application of advanced signal processing techniques in order to detect and classify the ionospheric abnormal signals during the geomagnetic and seismic activities. In recent years, powerful algorithms have been developed that can use logarithmically increasing data and powerful hardware to process them. Algorithms for Machine Learning (ML) are concerned with the design and development of algorithms that allow computers to learn based on different forms of data, including sensor data or databases. The goal of ML algorithms is to enable computers to recognize the complex patterns and arrive at informed judgments using data. In disciplines including statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control, and theoretical computer science, ML is so widely used. To train computers to carry out tasks for which there is not a totally suitable algorithm, ML employs a range of techniques. One strategy is to declare some of the right replies as valid when there are numerous possible responses. The algorithm(s) that the computer uses to discover the right answers can then be developed using this as training data. In order to analyze this data intelligently and develop relevant real-world applications, the ML algorithms are especially applied to very wide areas such as image processing with high-dimensional data [42], large cluster classification [43], finance [44], media [45], healthcare [46] and email/spam filtering [47]..

In this study, K-Nearest Neighbors (K-NN) algorithm, one of the ML techniques, is used to detect and classify the ionospheric disturbances during five strong earthquakes with magnitudes Mw greater than 5.6 and geomagnetic activity. The K-NN algorithm is a widely used algorithm because it is straightforward and simple to use, powerful and convenient to learn, and a non-parametric technique. The K-NN algorithm is a preferred option for many applications because it has the potential to analyze complex data patterns and is a simple and fast running algorithm. GPS-TEC obtained from International GNSS Service (IGS) centers as IONOLAB-TEC outputs with 2.5 min resolution constitutes the data set of this study. The K-NN algorithm is performed on GPS-TEC obtained for two geomagnetically quiet days, one geomagnetically disturbed day, five and three days before the each earthquake and the earthquake day. The data used in the study and the K-NN algorithm are presented in Sects. 2 and 3, respectively. The results are given in Sect. 4.

2 The Data

The K-NN algorithm is applied to TEC estimates obtained for two geomagnetically quiet days, one geomagnetically disturbed day, five and three days before the each earthquake and the earthquake day, respectively. TEC estimates are obtained from fifteen IGS-GPS stations listed in Table 2. Earthquake day's information are obtained from United States Geological Survey [48]. The chosen five earthquakes between 1999 and 2016 are listed in Table 1. The variations of Kp, Ap and Dst indices are given for the chosen disturbed days in Fig. 1.

TEC estimates used as the data source of this study are obtained from dual band GNSS (Global Navigation Satellite Systems) receivers. TEC data are estimated as IONOLAB-TEC

Table 1 The geographic locations, the date, the time, the magnitudes and the depths of the chosen earthquakes chosen for the study

Earthquakes	Date (DD MM YYYY)	Time (UTC) (hh:mm)	Latitude	Longitude	Magnitude Mw (Richter)	Depth (km)
E1	17 Aug 1999	00:02	40.74° N	29.86° E	7.4	17
E2	12 May 2008	06:28	31.02° N	103.36° E	7.9	10
E3	13 Jun 2008	23:43	39.03° N	140.88° E	6.9	7.8
E4	24 Aug 2016	01:36	42.71° N	13.17° E	6.2	10
E5	24 Aug 2016	02:33	42.79° N	13.15° E	5.6	10

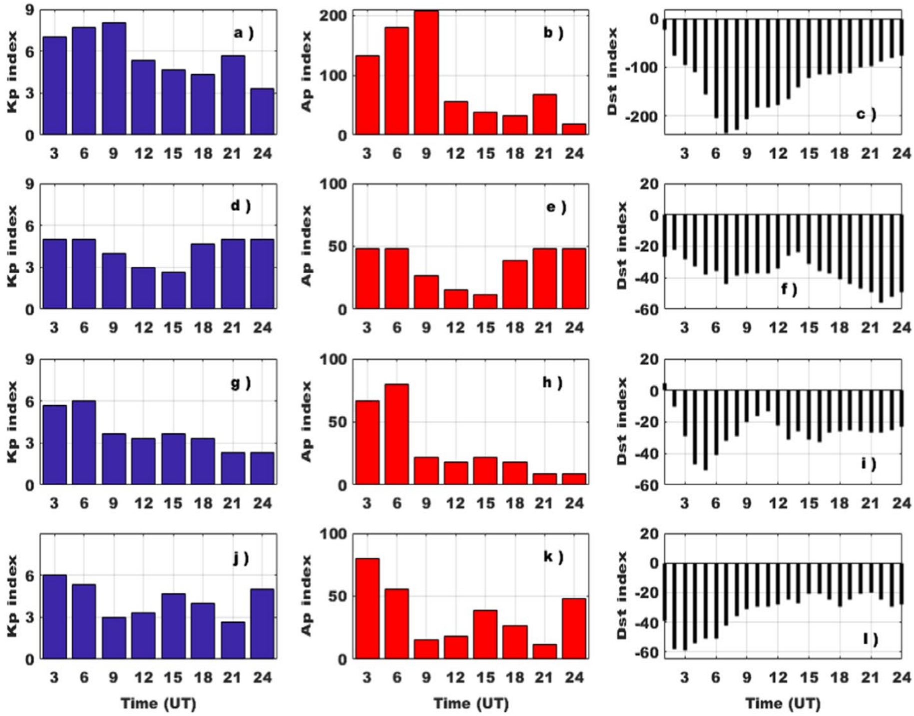


Fig. 1 The values of **a** Kp index, **b** Ap index and **c** Dst index for October 22, 1999; **d** Kp index, **e** Ap index and **f** Dst index for March 27, 2008; **g** Kp index, **h** Ap index and **i** Dst index for September 04, 2008 and **j** Kp index, k) Ap index and **l** Dst index for September 02, 2016

Table 2 The geographic locations and IDs of the GPS receiving stations used in the study

GPS Station	Receiver ID	Latitude	Longitude
Ankara, Turkey	ankr	39.69° N	32.75° E
Gebze, Turkey	tubi	40.59° N	29.45° E
Sofia, Bulgaria	sofi	42.55° N	23.39° E
Kunming, China	kunm	24.88° N	102.79° E
Lhasa, China	lhaz	29.49° N	91.10° E
Sheshan, China	shao	30.92° N	121.20° E
Shaanxi, China	xian	34.18° N	109.22° E
Mizusawa, Japan	mizu	38.94° N	141.13° E
Mitaka, Japan	mtka	35.49° N	139.56° E
Tsukuba, Japan	tskb	35.92° N	140.08° E
Usuda, Japan	usud	35.95° N	138.36° E
Genova, Italy	geno	44.22° N	08.92° E
Torino, Italy	ieng	44.82° N	07.63° E
Firenze, Italy	igmi	43.60° N	11.21° E
Matera, Italy	mat1	40.45° N	16.70° E

obtained with the Regularized Estimation (Reg-Est) algorithm. The Reg-Est algorithm is a powerful algorithm that can estimate the GPS-TEC on a single station in the direction of the local zenith as given in [14] and [49]. With the Reg-Est algorithm, GPS-TEC can be estimated both as IONOLAB-STEC and IONOLAB-TEC for any geographic coordinate for any days of ionosphere [50]. The ephemeris and IONospheric Exchange (IONEX) records including Differential Code Biases (DCB) and Global Ionospheric Maps (GIM) are obtained from Crustal Dynamics Data Information System (CDDIS), NASA's Archive Space Geodesy Data. In cases where Receiver DCBs are not available, IONEX files are computed using IONOLAB-BIAS algorithm given in [51]. The IONOLAB-STEC and IONOLAB-TEC can also be accessed from IONOLAB [15] as *.exe format [52]. In this study, 2.5 min time resolution IONOLAB-TEC data are estimated for each GPS station whose geographic coordinates are given in Table 2 [53].

The algorithm is applied into two groups. The ionosphere is relatively quiet both for Group I and II during the chosen earthquake periods. Kp and Ap indices have maximum values of 5 and 29, respectively. The Dst index reaches maximum -40 nT. The Kp and the Ap indices are downloaded from Space Weather Prediction Center of National Oceanic and Atmospheric Administration [54]. The Dst indices are achieved from World Data Center for Geomagnetism [55].

3 The K-Nearest Neighbors Algorithm

IONOLAB-TEC used in the study can be defined for day d of any u station as follows:

$$\boldsymbol{x}_{u;d} = \left[x_{u;d}(1) \ldots x_{u;d}(n) \ldots x_{u;d}(N)\right]^T, \tag{1}$$

where $x_u;d(n)$ is the nth TEC measurement, $x_u;d(N)$ is the total amount of TEC samples on day d and T is the transpose of the vector.

In many pattern classification studies, feature vectors are created using statistical features. In this study, different days of any station compose the feature matrices for each station from the TEC vectors defined by Eq. (1). For any station u, the feature matrix F_u can be defined as follows:

$$F_u = \left[\boldsymbol{x}_{u;d1}, \boldsymbol{x}_{u;d2}, \ldots, \boldsymbol{x}_{u;dm}\right]^{\boldsymbol{T}} \tag{2}$$

where, the vectors $\boldsymbol{x}_{u;d1}, \boldsymbol{x}_{u;d2}, \ldots, \boldsymbol{x}_{u;d\mathbf{m}}$ represent the daily IONOLAB-TEC vectors between the first and the last days of station u, respectively. For each earthquake, the feature matrix F_E combined with daily IONOLAB-TEC vectors of different stations u, v, y and z can be defined as follows:

$$F_E = \left[\boldsymbol{x}_{u;d1}, \ldots, \boldsymbol{x}_{u;dm}; \boldsymbol{x}_{v;d1}, \ldots, \boldsymbol{x}_{v;dm}; \boldsymbol{x}_{y;d1}, \ldots, \boldsymbol{x}_{y;d11}; \boldsymbol{x}_{z;d1}, \ldots, \boldsymbol{x}_{z;dm}\right]^{\boldsymbol{T}} \tag{3}$$

The algorithm is applied using two different groups of approaches. The first two columns of the feature matrices F are accepted as having a non-earthquake and non-geomagnetic activity status. The fourth column is accepted as geomagnetic disturbed status. The following five columns (columns 5–9) in Group I are accepted as having a possible precursor status, and the final column (column 10) is recognized as having an earthquake status. The columns between five and seven are designated as a possible precursor status in Group II. The last column (8) in Group II is designated as an earthquake status.

To determine which day the earthquake precursor signals appear is the basis for two techniques of the study.

Machine learning algorithms can be grouped into parametric and nonparametric models. The K-Nearest Neighbors (K-NN) algorithm belongs to a subcategory of non-parametric models defined as instance-based learning. Instance-based learning contains the nearest neighbors and locally weighted regression methods conjecturing the samples can be symbolized as points in the Euclidean space. The K-NN algorithm is an algorithm performed on solving both classification and regression algorithms and can be easily adapted to approximate continuous-valued target functions. The K-NN algorithm is quickly executed for small training datasets. The algorithm does not need any prior knowledge of the pattern of the data in the training set. If a new training set is added to the existing pattern in the algorithm, no retraining is required. One of the critical points for the performance of the K-NN algorithm is how to measure the closeness between samples [56, 57].

Euclidean distance is a highly preferred metric for continuous variables located on a straight line in Euclidean space. The K-NN algorithm classifies an unknown sample presented by some feature vector as a point in the feature space by calculating the distance between the points in the training data set. The algorithm specifies the class of the point between its k nearest neighbors [58]. When a new query example comes, its relationship with the previously stored examples is tested in order to appoint a target function value for the new example. In this study, the Euclidean distance between two days *da* and *db* of the feature matrix of any station u is defined as follows [59]:

$$d(F_u) = \sqrt{\sum_{n=1}^{k} \left(x_{u;da}(n) - x_{u;db}(n)\right)^2} \tag{4}$$

For each earthquake, Eq. (4) can be defined for the feature matrix F_E combined with daily IONOLAB-TEC vectors of different stations as follow:

$$d(F_E) = \sqrt{\sum_{n=1}^{k} \left(x_{u;d}(n) - x_{v;d}(n)\right)^2} \tag{5}$$

To estimate the value of the test data by weighted of its neighbors, the inverse Euclidean distance weight is computed as follows:

$$w_u = \frac{1}{F_u} \tag{6}$$

When the unbundled variables in the data set are in different units, it is crucial to normalize the variables before calculating the distance. This poses a problem, as most machine learning algorithms use the Euclidean distance between two data points in their calculations. So in this case, both properties must be done in the same range. In order to compensate this impact, it is necessary to get all properties to the same order of magnitude. This can be accepted as normalization. In this study, Min Max Scaler (M*MS*) function defined as follows is used for the normalization:

$$MMS = \frac{F_u(n) - min(F_u)}{max(F_u) - min(F_u)} \tag{7}$$

How the model is applied to data sets can be seen from the flow diagram given in Fig. 2.

After the training process, the classification success is checked with the test data. Four different evaluation criteria, the Accuracy (A), Precision (P), Recall (RC) and $F1-score$, the parameters are used to evaluate the model performance. The formulas for these evaluation metrics are defined as follows:

$$A = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$RC = \frac{TP}{TP + FN} \tag{10}$$

$$F1 - score = 2x\frac{PxRC}{P + RC} \tag{11}$$

The higher A value is obtained where the most common evaluation metric and the better the classification performance is calculated. P and RC are used to evaluate the accuracy of the classifier. The small RC value shows that the classifier is aggrieved from many False Negatives (FN) and the small P value shows that the classifier is aggrieved from a large number of False Positives (FP). While True Positive (TP) refers to the correctly classified diseased cases, False Negative (FN) refers to the cases with the wrongly classified disease. Similarly, True Negative (TN) refers to correctly classified healthy cases, while FP indicates incorrectly classified healthy cases. The $F1-score$ is used to prevent incorrect model selection in the data set that is not evenly distributed.

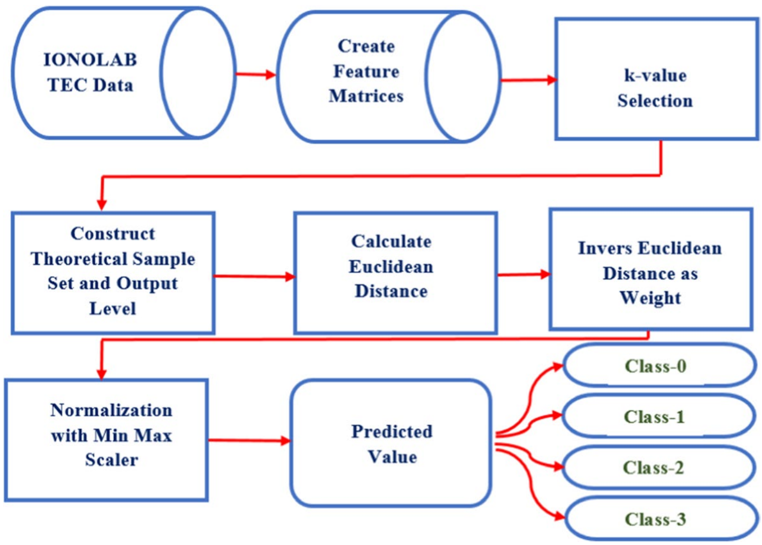


Fig. 2 The block diagram of the K-NN algorithm used in the study

4 Results and Discussion

In this study, the K-NN algorithm defined in Sect. 3 is performed on IONOLAB-TEC estimates in order to classify the ionospheric disturbances during five different earthquakes periods with magnitudes Mw between 5.6 and 7.9 and geomagnetic activity periods. For each earthquake, IONOLAB-TEC estimates are obtained from at least three stations. The stations are chosen from those closest to the earthquake center and data available. The lack of coarse data for chosen stations and days in the study is the cause of the missing values in the chosen periods. The geographic coordinates of these GPS stations given in Table 2, distances to the epicenters and the coordinates of the earthquakes given in Table 1 are demonstrated in Fig. 3.

Investigation periods for each earthquake are chosen into two groups. In Group I, five days before the each earthquake are chosen as a possible earthquake precursors. In Group II, three days before the each earthquake are chosen as the same status. For both two groups of application, two geomagnetically quiet days and one geomagnetically disturbed day are chosen as non-earthquake and non-geomagnetic activity and geomagnetic activity, respectively. In Fig. 4, the IONOLAB-TEC estimates for the closest stations to the epicenters are given for one quiet day (Fig. 4a, d, g and j) for E1, E2, E3, E4 and E5, respectively. The chosen geomagnetic disturbed days are represented in Fig. 4b, e, h and k. The red lines in Fig. 4e, f, i and l demonstrate the five days before and the earthquake days of E1, E2, E3, E4 and E5, respectively. The distance between the stations to the earthquake epicenters vary from 38 to 651 km. In Fig. 4, it is observed that TEC has minimum value one day before E1, three and five days before E2, one day before E3 and two days before E4 and E5. Also,

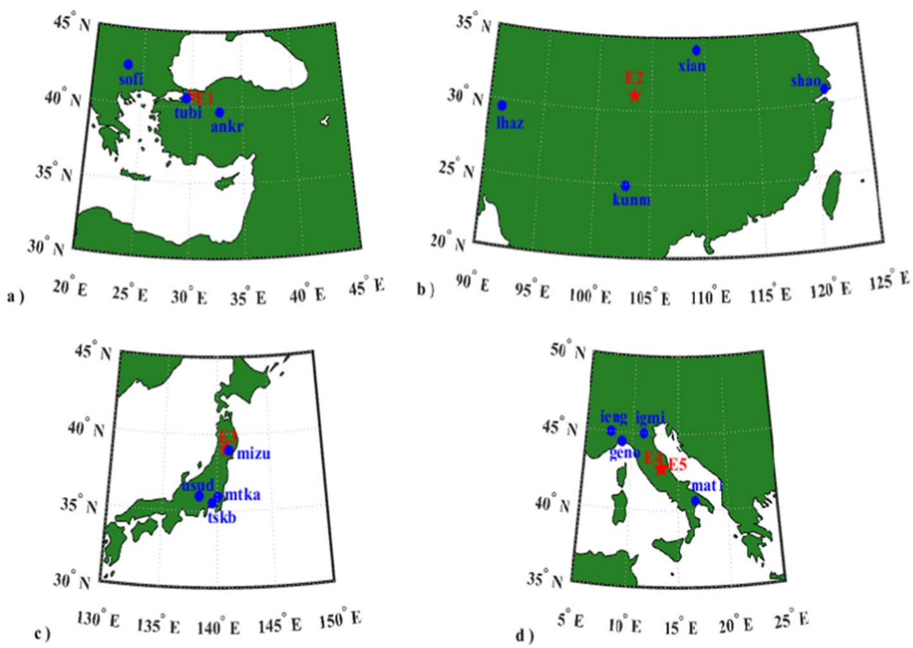


Fig. 3 Locations of reference stations given in Table 2 shown with blue circles and earthquakes shown with red stars for: **a** E1, **b** E2, **c** E3 and **d** E4 and E5 given in Table 1

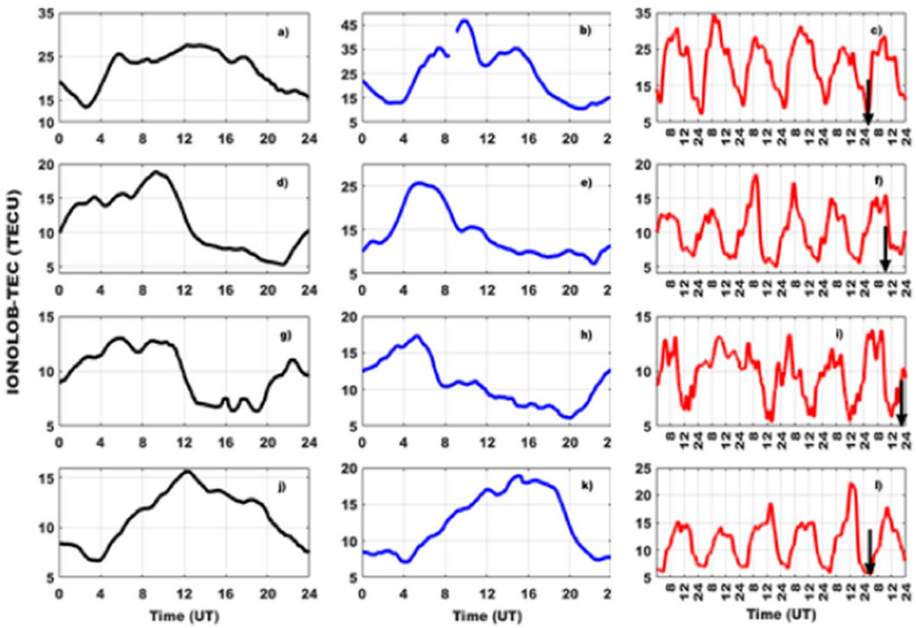


Fig. 4 IONOLAB-TEC values for tubi: **a** on August 07, 1999, **b** on October 22, 1999 and **c** between August 12 and 17, 1999; for xian: **d** on May 02, 2008, **e** on March 27, 2008 and **f** between May 02 and 12, 2008; for mizu: **g** on June 03, 2008, **h** on September 04, 2008 and **i** between June 03 and 13, 2008 and for igmi: **j** on August 24, 2016, **k** on September 02, 2016 and **l** between August 14 and 24, 2016. Black and blue solid lines represent IONOLAB-TEC values obtained for the geomagnetically quiet and disturbed days, respectively. Red solid lines represent the IONOLAB-TEC values obtained for the six-day earthquake period including five days before the earthquake and the day of the earthquake

TEC has maximum value four days before E1 and one day before E4 and E5. Three and four days prior to the earthquake, the normal TEC corruption occurs during the E3 period. In all geomagnetic disturbed days, TEC values are much greater than on all quiet days in Fig. 4.

In Group I, two days are marked as the non-earthquake and non-geomagnetic activity as 0-class. One geomagnetic disturbed day (given in Fig. 4) is marked as 1-class. Five days just before the each earthquake day are marked as the earthquake precursor status as 2-class. The earthquake day is assigned as earthquake status as 3-class. In Group I, similar to Group I, two non-earthquake and non-geomagnetic activity days and one disturbed day are as 0-class and 1-class, respectively. Three days just before the each earthquake day are marked as the earthquake precursor status as 2-class. The earthquake day is assigned as earthquake status as 3-class. 75% and 25% of the data consist of the feature matrices are used for training and test, respectively. In this distinction, the samples are chosen randomly. After the training process, the classification success is checked with the test data. Then, the K-NN algorithm given in Sect. 3 is applied to the feature matrices during the training and the test processes.

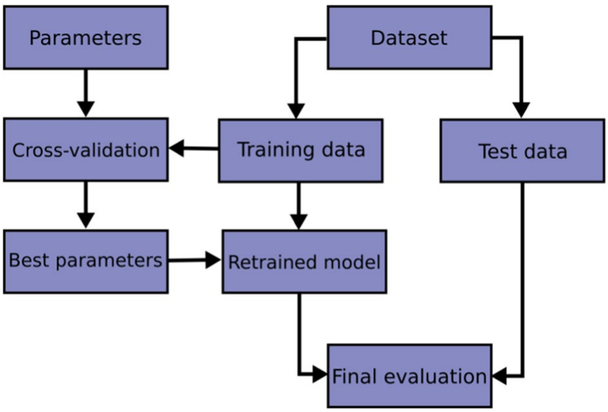
A methodological error occurs when a forecasting function's parameters are obtained and then evaluated on the same data: a model that could only repeat the samples' tags that it has just seen would have an ideal score, but it would be unable to predict anything useful on data that has not yet been seen. The term "overfitting" describes this scenario. In order to avoid it, it is customary to reserve a subset of the available data as a test set (x_{test}, y_{test})

while conducting a (supervised) Machine Learning experiment. Note that even in commercial settings, Machine Learning often begins with an experiment, hence the term "trial" is not limited to use in academic contexts only. In Fig. 5, a flow chart is given that shows how the standard model training procedure for cross-validation is carried out.

In this study, the K-NN algorithm given in Sect. 3 is performed on the IONOLAB-TEC estimates obtained during five earthquakes periods with magnitudes Mw between 5.6 and 7.9, geomagnetic storms and quiet days. IONOLAB-TEC estimates are normalized using Eq. (7). The data sets for each earthquake are consists of a total of 5184 samples of 9 days for Group I and 4632 of 7 days for Group II at 2.5 min resolution for one station. The feature matrix given by Eq. (5) for each earthquake consists of three station data for the E1 earthquake and four station data for the E2, E3 and E4 earthquakes. All stations located around the epicenter of the earthquakes and providing data are used within the scope of the study. As a result of the application of the Eq. (5), the confusion matrices obtained for each earthquake are given in Figs. 6 and 7 for Group I and Group II, respectively. In the confusion matrices where the true classes and predicted classes are shown, the four status are given 0, 1, 2 and 3, respectively. When Figs. 6 and 7 are both investigated, it is observed that the algorithm has the capability to distinguish the classification of four status in the predictive and the true classes. How many of the test classes are accurately predicted when compared to the true classes reveals the algorithm's overall classification Accuracy. When the confusion matrices are compared with each other, it is observed that the classification success performance is generally high. The counts of the TP classes for all three status are greater than 96%. The earthquake with the lowest rate of misclassification is observed at status 1 for E2 for both Group I and Group 2 (Figs. 6a and 7a). In Group 1, 267 samples are classified as non-earthquake and non geomagnetic activity (98.8%), all 138 samples are classified as geomagnetic activity (100%), all 751 samples are classified as possible earthquake precursor (100%) and all 137 samples are classified as earthquake (100%) for E2. In Group II, 1 and 3 classes are classified with 100% Accuracy while 0 and 2 classes are classified with 99.3% and 99.7 for E2, respectively.

Accuracy (A, Eq. (8)), Precision (P, Eq. (9)), Recall (RC, Eq. (10)) and $F1-score$ (Eq. (11)) performance metrics are computed for each earthquake using the confusion matrices given in Figs. 6 and 7 to evaluate the results. The metrics$A$,P, RC and $F1-score$ values are comparatively given in Table 3 both for Group I and II for 2-class. It is observed that there is a positive correlation between the A and the magnitude of the earthquake. The A values

Fig. 5 Flow chart for the standard model training procedure for cross-validation



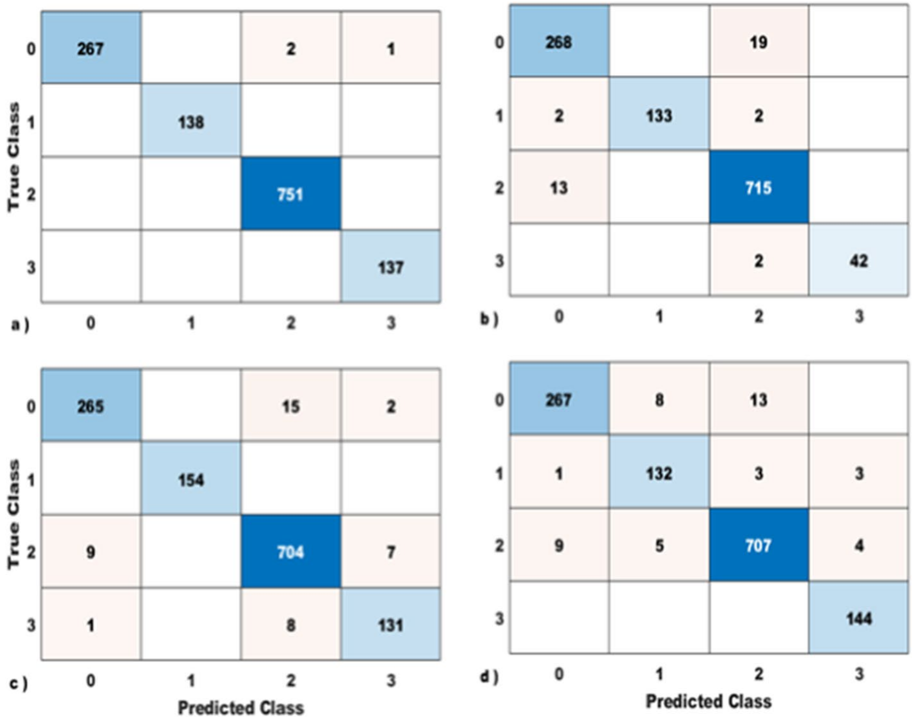


Fig. 6 The confusion matrices for Group I in order of magnitude from greater to smaller for: **a** E2, **b** E1, **c** E3 and **d** E4 and E5

increase directly related to the magnitude of the earthquakes. *A* values for Group I are greater than those for Group II when A values for Groups I and II are compared to one another. All *A* values for Group I and II are over 96% for all four earthquakes. Small differences are observed between classification accuracies of each region in the analysis results. In Machine Learning, *P* and *RC* values should be 1 for good classification. For both Group I and Group II, a positive correlation is observed between *A* values and earthquake magnitude. *A* values increase depending on the magnitude of the earthquake. In this study, *P* values for two groups are at least 95% and above. The $F1-score$ values, which is the weighted average of the *P* and*RC*, are above 96% for both two groups. As shown in Table 3, the greater $F1-score$ values indicate that the classifier has the smaller *FP* and *FN* values. It is observed that the proposed method has a good performance to be able to classify the possible earthquake precursor signals in the ionosphere based on the *A* and $F1-score$ values in the study.

As a result of the application of the Eq. (4), *A* values are computed for each stations within five different distance categories given in Table 4. The Accuracy values for each station are comparatively given for class-2 both for Group I and II in Tables 5 and 6. It is observed that the *A* values decrease as the distance of the station to the epicenter increases for both Group I and II. It is observed that the *A* values decrease as the distance of the station to the epicenter increases for both Group I and II. While the accuracy for stations in the category C1 for Group I is about 60%, the *A* is about 50% for Group II. For Group II, the *A* is about 50% for the stations in C1 and C2, but less than 50% for the stations in C3, C4 and C5. If 50% for *A* values is accepted as the threshold for both two groups, *A* values

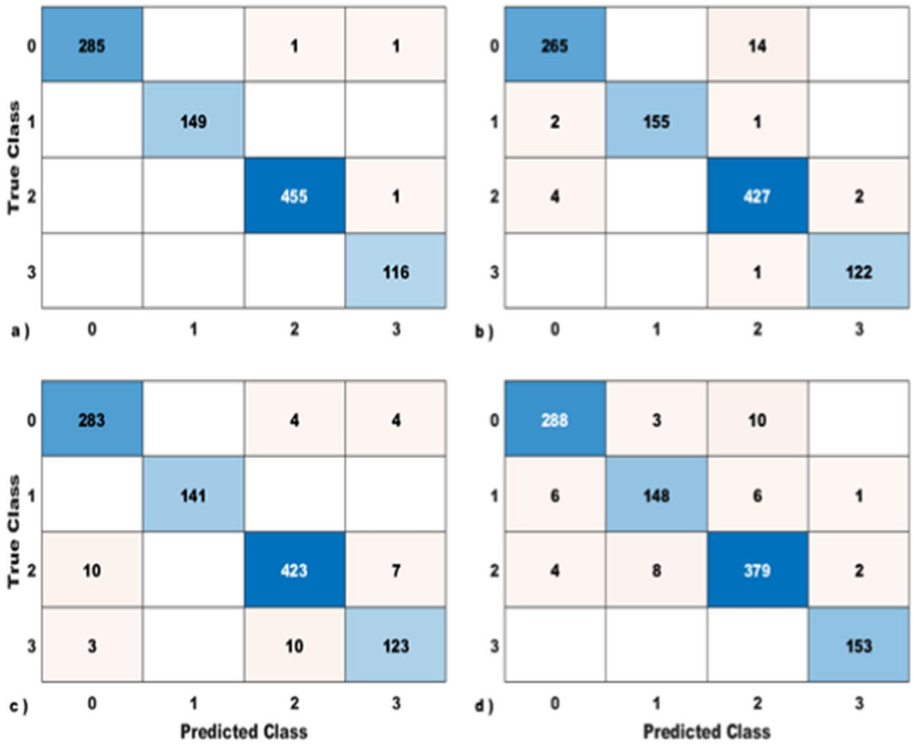


Fig. 7 The confusion matrices for Group II in order of magnitude from greater to smaller for: **a** E2, **b** E1, **c** E3 and **d** E4 and E5

Table 3 Accuracy (A), Precision (P), Recall (RC) and $F1 - score$ values of the all earthquakes arranged in order of magnitude from highest to lowest

	$A(\%)$		$P(\%)$		$RC(\%)$		$F1 - score(\%)$	
	Group I	Group II	Group I	Group II	Group I	Group II	Group I	Group II
E2	99.8	99.7	99.7	99.8	100	99.8	99.8	99.8
E1	96.8	97.6	96.9	96.4	98.2	98.6	97.5	97.5
E3	96.8	96.2	96.8	96.8	97.7	96.1	97.2	96.4
E4	96.5	96.0	97.7	95.9	97.5	96.4	97.5	96.1

Table 4 The categories grouped according to the distance of the stations from the epicenters

Category	Distance to Epicenter (km)	Stations
C1	0–50	mizu, tubi
C2	50–300	igmi, ankr
C3	300–500	tskb, geno, mat1, usud, mtka, ieng
C4	500–1000	sofi, kunm, xian
C5	>1000	lhaz, shao

Table 5 Accuracy (A) values of the all stations ordered according to their distance to the epicenters from the nearest to the farthest for Group I

Categories	E2		E1		E3		E4	
	Station	$A(\%)$	Station	$A(\%$	Station	$A(\%)$	Station	$A(\%)$
C1	–	–	tubi	57.8	mizu	60.7	–	–
C2	–	–	ankr	52.4	–	–	igmi	51.3
C3	–	–	–	–	tskb	51.3	geno	50.7
	–	–	–	–	usud	50.8	mat1	50.7
	–	–	–	–	mtka	50.4	ieng	49.6
C4	xian	49.5	sofi	47.4	–	–	–	–
	kunm	49.0	–	–	–	–	–	–
C5	lhaz	42.1	–	–	–	–	–	–
	shao	41.9	–	–	–	–	–	–

Table 6 Accuracy (A) values of the all stations ordered according to their distance to the epicenters from the nearest to the farthest for Group II

Categories	E2		E1		E3		E4	
	Station	$A(\%)$	Station	$A(\%)$	Station	$A(\%)$	Station	$A(\%)$
C1	–	–	tubi	48.7	mizu	49.0	–	–
C2	–	–	ankr	47.4	–	–	igmi	46.0
C3	–	–	–	–	tskb	43.3	geno	43.9
	–	–	–	–	usud	43.1	mat1	43.8
	–	–	–	–	mtka	43.1	ieng	42.4
C4	xian	42.7	sofi	42.8	–	–	–	–
	kunm	42.9	–	–	–	–	–	–
C5	lhaz	41.8	–	–	–	–	–	–
	shao	40.8	–	–	–	–	–	–

are greater than 50% for stations up to 272 km from the epicenters as they are smaller than 50% for stations up to 500 km from the epicenters.

In this study, the reason for taking the harmonic mean instead of a simple mean while obtaining the experimental results is that extreme cases should not be ignored. Assuming the F1-score is simply averaged, a model with a P value of 1 and a RC value of 0 will have an F1-score of 0.5. This would be misleading. The main reason for using the F1-score parameter to evaluate model performance is to prevent a misleading result in an unevenly distributed (unbalanced) data set.

Classifying the precursor signals in the ionosphere related to earthquakes can create new opportunities for early warning systems. Accurately forecasting the occurrence of the earthquakes has always been difficult since earthquakes are intricate geological phenomena. According to this study, tracking specific ionospheric anomalies may yield important insights prior to earthquake occurrences. Comprehending precursor signals in the ionosphere aids in the monitoring of ionospheric disturbances and anomalies. Not only can monitoring these signals help with earthquake prediction, but it also sheds light on space weather events. Enhancing ionospheric monitoring can assist reduce the effects of space weather events on GPS, satellite communication, and other technologies, improving their performance and dependability. This study can motivate multi-disciplinary collaboration between seismologists and space weather researchers. These disciplines have historically worked independently, but the results in this study can point to the possibility that improvements in both might come from a better comprehension of ionospheric processes. Through this collaboration, it may be possible to create early warning systems that are more reliable and to comprehend the intricate relationships that exist between Earth's geophysical and space environments.

5 Conclusion

In this study, the ionospheric disturbances due to the seismic and geomagnetic activity are classified with the K-Nearest Neighbors (K-NN) algorithm by using GPS-TEC. TEC data is estimated for 15 IGS stations during five earthquakes with magnitude between Mw 7.9 and 5.6, one geomagnetically disturbed day and two quiet days with no seismic and geomagnetic activity. The results are obtained into two application groups. In two groups, the three non-earthquake and non-geomagnetic activity days are marked as the non-earthquake status as 0-class. One geomagnetic disturbed day is assigned as 1-class. In Group I, five days before the each earthquake are accepted as the earthquake precursor status as 2-class. In Group II, three days before the each earthquake are accepted as the earthquake precursor status as 2-class. The earthquake days are assigned as earthquake status as 3-class both for Group I and II. 75% and 25% of the data sets are used for training and test, respectively. The K-NN algorithm is applied to the trained and tested data for all earthquakes and stations. Four different metrics, the Accuracy (A), Precision (P), Recall (RC) and $F1 - score$, are calculated using confusion matrices to evaluate the performance of the model. Out of 5184 test samples, it is observed that 5055 are classed as true (97.5%), while 129 are classified as false (2.5%) for Group I. Out of 4017 test samples, 3912 and 105 samples are classified as true (97.4%) and false (2.6%) for Group II, respectively. The Accuracy values for the possible earthquake class rise inversely with the distance of the stations to the epicenters and are correlated with the magnitude of the earthquakes.

The results of this study may potentially be useful in the development of technologies for better monitoring both space weather and pre-earthquake process and forecasting. This could result in the improvement of current monitoring systems and the creation of new tools that have a greater degree of precision in detecting ionospheric precursors. In conclusion, the results of this study on classification ionospheric disturbances due to geomagnetic and seismic activity have significant consequences for the fields of space weather and seismology researches. Enhancing ionospheric monitoring and improving earthquake prediction are potential benefits that highlight the significance of ongoing research and cooperation in these areas for the advancement of knowledge in science and real-world applications.

Acknowledgements The authors wish to thank Prof. Dr. Feza Arikan and IONOLAB group for their outstanding efforts on IONOLAB-BIAS and IONOLAB-TEC Algorithm.

Author Contributions CB, SK, Faruk Erken and Ali Cinar made all calculations, created the algorithm and wrote the main manuscript. All authors reviewed the manuscript.

Funding The authors declare that no funds, grants, or other supports were received during the preparation of this manuscript.

Data Availability The corresponding author may provide the data and material used in the manuscript subjected to reasonable request.

Declarations

Conflict of interest The authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Santis, A. D., Marchetti, D., Pavón-Carrasco, F. J., et al. (2019). Precursory worldwide signatures of earthquake occurrences on Swarm satellite data. *Nature Scientific Reports, 9*(20287), 1–13.
2. Akyol, A., Arikan, O., & Arikan, F. (2020). A machine learning-based detection of earthquake precursors using ionospheric data. *Radio Science, 55*(11), 1–21.
3. Liu, J., Wang, W., Zhang, X., Wang, Z., & Zhou, C. (2022). Ionospheric total electron content anomaly possibly associated with the April 4, 2010 Mw7.2 Baja California earthquake. *Advances in Space Research, 69*(5), 2126–2141.
4. Akhoondzadeh, M., Santis, A. D., Marchetti, D., & Wang, T. (2022). Developing a deep learning-based detector of magnetic, Ne, Te and TEC anomalies from swarm satellites: The case of Mw 7.1 2021 Japan Earthquake. *Remote Sensing, 14*(7), 1–22.
5. Petrescu, L., & Moldovan, I. A. (2022). Prospective neural network model for seismic precursory signal detection in geomagnetic field records. *Machine Learning and Knowledge Extraction, 4*(4), 912–923.
6. Gurbuz, G., Aktug, B., Jin, S., & Kutoglu, S. H. (2020). A GNSS-based near real time automatic Earth Crust and Atmosphere Monitoring Service for Turkey. *Advances in Space Research, 66*(12), 2854–2864.
7. Pulinets, S. A. (2004). Ionospheric precursors of earthquakes: Recent advances in theory and practical applications. *Terrestrial Atmospheric and Oceanic Sciences, 15*(3), 413–435.
8. Pulinets, S. A., Gaivoronska, T. B., Contreras, A. L., & Ciraolo, L. (2004). Correlation analysis technique revealing ionospheric precursors of earthquakes. *Natural Hazards and Earth System Sciences, 4*, 697–702.
9. 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 4248–4263.
10. Rishbeth, H., & Garriott, O. K. (1969). *Introduction to ionospheric physics*. Academic Press.
11. Karatay, S. (2020). Detection of the ionospheric disturbances on GPS-TEC using differential rate Of TEC (DROT) algorithm. *Advances in Space Research, 65*(10), 2372–2390.
12. Karatay, S. (2020). Estimation of frequency and duration of ionospheric disturbances over Turkey with IONOLAB-FFT algorithm. *Journal of Geodesy, 94*(89), 1–24.
13. Erken, F., Karatay, S., & Cinar, A. (2019). Spatio-temporal prediction of ionospheric total electron content using an adaptive data fusion technique. *Geomagnetism and Aeronomy, 59*, 971–979.
14. Arikan, F., Erol, C., & Arikan, O. (2003). Regularized estimation of vertical total electron content from Global Positioning System data. *Space Physics, 108*(A12), 1–20.
15. 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), 297–305.
16. Laštovička, J. (1996). Effects of geomagnetic storms in the lower ionosphere, middle atmosphere and troposphere. *Journal of Atmospheric and Terrestrial Physics, 58*(7), 831–843.
17. Chen, Y., Liu, L., Le, H., Zhang, H., & Zhang, R. (2022). Responding trends of ionospheric F2-layer to weaker geomagnetic activities. *Journal of Space Weather and Space Climate, 12*(6), 1–12.
18. Li, H., Wang, J.-S., Chen, Z., Xie, L., Li, F., & Zheng, T. (2020). The contribution of geomagnetic activity to ionospheric foF2 trends at different phases of the solar cycle by SWM. *Atmosphere, 11*(6), 1–12.
19. Pulinets, S., & Ouzounov, D. (2011). Lithosphere–Atmosphere–Ionosphere Coupling (LAIC) model—An unified concept for earthquake precursors validation. *Journal of Asian Earth Sciences, 41*(4–5), 371–382.

20. Carbone, V., Piersanti, M., Materassi, M., Battiston, R., Lepreti, F., & Ubertini, P. (2021). A mathematical model of lithosphere–atmosphere coupling for seismic events. *Nature Scientific Reports, 11*, 1–12.
21. Bolt, B. (1964). Seismic air waves from the great 1964 Alaskan earthquake. *Nature, 202*, 1095–1096.
22. Donn, W. L., & Posmentier, E. S. (1964). Ground-coupled air waves from the Great Alaskan Earthquake. *Journal of Geophysical Research, 69*(24), 5357–5361.
23. Davies, K., & Baker, D. M. (1965). Ionospheric effects observed around the time of the Alaskan earthquake of March 28, 1964. *Journal of Geophysical Research, 70*(9), 2251–2253.
24. Leonard, R. S., & Barnes Jr, R. A. (1965). Observation of ionospheric disturbances following the Alaska earthquake. *Journal of Geophysical Research, 70*(5), 1250–1253.
25. Row, R. V. (1966). Evidence of long-period acoustic-gravity waves launched into the F region by the Alaskan earthquake of March 28, 1964. *Journal of Geophysical Research, 71*(1), 343–345.
26. Hirshberg, J., Currie, R. G., & Breiner, S. (1967). Long period geomagnetic fluctuations after the 1964 Alaskan earthquake. *Earth and Planetary Science Letters, 3*, 426–428.
27. Yuen, P. C., Weaver, P. F., Suzuki, R. K., & Furumoto, A. S. (1969). Continuous, traveling coupling between seismic waves and the ionosphere evident in May 1968 Japan earthquake data. *Journal of Geophysical Research, 74*(9), 2256–2264.
28. Weaver, P. F., Yuen, P. C., Prolss, G. W., & Furumoto, A. S. (1970). Acoustic coupling into the ionosphere from seismic waves of the earthquake at Kurile Islands on August 11, 1969. *Nature, 226*, 1239–1241.
29. Antsilevich, M. G. (1971). The influence of Tashkent earthquake on the earth's magnetic field and the ionosphere, Tashkent earthquake 26 April 1966. In *FAN Publishing House*, pp. 187–188.
30. Datchenko, E., Ulomov, V., & Chernyshova, C. (1973). Electron density anomalies as the possible precursor of Tashkent earthquake. *Academy of Sciences, 12*, 30–32.
31. Larkina, V. I., Migulin, V. V., Molchanov, O. A., Kharkov, I. P., Inchin, A. S., & Schvetcova, V. B. (1989). Some statistical results on very low frequency radiowave emissions in the upper ionosphere over earthquake zones. *Physics of the Earth and Planetary Interiors, 57*(1–2), 100–109.
32. Liu, J. Y., Chen, Y. I., Pulinets, S. A., Tsai, Y. B., & Chuo, Y. J. (2000). Seismo-ionospheric signatures prior to M≥6.0 Taiwan earthquakes. *Geophysical Research Letters, 27*(19), 3113–3116.
33. Pulinets, S. A., Contreras, A. L., Bisiacchi-Giraldi, G., & Ciraolo, L. (2005). Total eletron content variations in the ionosphere before the Colima, Mexico, earthquake of 21 January 2003. *Geofísica Internacional, 44*(4), 369–377.
34. Pulinets, S., Kotsarenko, A., Ciraolo, L., & Pulinets, I. A. (2007). Special case of ionospheric day-to-day variability associated with earthquake preparation. *Advances in Space Research, 39*(5), 970–977.
35. Karatay, S., Arikan, F., & Arikan, O. (2010). Investigation of total electron content variability due to seismic and geomagnetic disturbances in the ionosphere. *Radio Science, 4*(5), 1–12.
36. Pulinets, S. A., Khegal, V. V., Boyarchuk, K. A., & Lomonosov, A. M. (1998). The atmospheric electric field as a source of variability in the ionosphere. *Physics-Uspekhi, 41*(5), 515–522.
37. Liu, J. Y., et al. (2004). Ionospheric foF2 and TEC anomalous days associated with M >= 5.0 earthquakes in Taiwan during 1997–1999. *Terrestrial Atmospheric and Oceanic Sciences, 15*(3), 371–383.
38. Liu, J. Y., Chen, Y. I., Chen, C. H., & Hattori, K. (2010). Temporal and spatial precursors in the ionospheric global positioning system (GPS) total electron content observed before the 26 December 2004 M9.3 Sumatra-Andaman Earthquake. *Journal of Geophysical Research Space Physics, 115*(A9), 1–13.
39. Kouris, S., Polimeris, K., & Cander, L. R. (2006). Specifications of TEC variability. *Advances in Space Research, 37*(5), 983–1004.
40. Akhoondzadeh, M. (2016). Decision Tree, Bagging and Random Forest methods detect TEC seismo-ionospheric anomalies around the time of the Chile, (Mw = 8.8) earthquake of 27 February 2010. *Advances in Space Research, 57*(12), 2464–2469.
41. Davidenko, D. V., & Pulinets, S. A. (2019). Deterministic variability of the ionosphere on the eve of strong (M ≥ 6) earthquakes in the regions of Greece and Italy according to long-term measurements data. *Geomagnetism and Aeronomy, 59*, 493–508.
42. Budak, C., Turk, M., & Toprak, A. (2016). Removal of impulse noise in digital images with na"ıv al of impulse noise in digital images with naııve Bayes. *Turkish Journal of Electrical Engineering and Computer Sciences, 24*(4), 2717–2729.
43. Albayrak, A. (2022). Classification of analyzable metaphase images using transfer learning and fine tuning. *Medical & Biological Engineering & Computing, 60*, 239–248.
44. Sarea, A. M., Subramanian, S., & Alareeni, B. (2021). Web-based financial disclosures by using machine learning analysis: Evidence from Bahrain. In *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success,* (pp. 357–371) Springer.

45. Kim J. H., Kim B. S., & Savarese S. (2012). Comparing image classification methods: K-nearest-neighbor and support-vector-machines. In *6th WSEAS international conference on Computer Engineering and Applications*, Harvard Cambridge.
46. Budak, C., & Mencik, A. (2022). Detection of ring cell cancer in histopathological images with region of interest determined by SLIC superpixels method. *Neural Computing and Applications, 34*, 13499–13512.
47. Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to Spam filtering. *Expert Systems with Applications, 36*(7), 10206–10222.
48. USGS, United States Geological Survey Earthquake Hazards Program. Available: https://earthquake.usgs.gov/.
49. Arikan, F., Erol, C., & Arikan, O. (2004). Regularized estimation of vertical total electron content from GPS data for a desired time period. *Radio Science, 39*(6), 1–10.
50. Nayir, H., Arikan, F., Arikan, O., & Erol, C. (2007). Total electron content estimation with reg-est. *Journal of Geophysical Research Space Physics, 112*(A11), 1–11.
51. Arikan, F., Nayir, H., Sezen, U., & Arikan, O. (2008). Estimation of single station interfrequency receiver bias using GPS-TEC. *Radio Science, 43*(4), 1–13.
52. Arikan F., Sezen U., Toker C., Artuner H., Bulu G., Demir U., Erdem E., Arikan O., Tuna H., Gulyaeva T. L., Karatay S., & Mosna Z., (2016). Space weather studies of IONOLAB group. In *URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, Seul.
53. Bureau I. C. International GNSS Service, NASA Jet Propulsion Laboratory California Institute of Technology. Available: https://igs.org/network/. Retrieved 2022.
54. NOAA, National Oceanic and Atmospheric Administration. Available: ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/.
55. WDC, World Data Center for Geomagnetism, Kyoto, [Online]. https://wdc.kugi.kyoto-u.ac.jp/. Retrieved 20 Feb 2023.
56. Tasci E., & Onan, A. (2017) K-En Yakın Komşu Algoritması Parametrelerinin. In *Akademik Bilisim Conference*, Aydin Turkey.
57. Raschka, S. (1969). *Python machine learning*. Packt Publishing.
58. Hu, L.-Y., Huang, M.-W., Ke, S.-W., & Tsai, C.-F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus, cilt 5, no. 1304, pp. 1–9.
59. Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. In *International conference on ıntelligent computing and control systems (ICCS)*, Madurai India.

Cafer Budak received his undergraduate, graduate and Ph.D. from Dicle University, Dicle University and Firat University in 1998, 2008 and 2014, respectively. He is currently an associate professor at Department of Electrical and Electronics Engineering, Dicle University. His current research interests are Telecommunication, Signal Processing and Deep Learning. He is Vice Chairman of the Commission H of URSI Turkey.

Secil Karatay is currently an associate professor with the Department of Electrical and Electronics Engineering, Kastamonu University. She received her undergraduate, graduate and Ph.D. education in 2001, 2005 and 2010 from Firat University, respectively. Her current research interests are in Digital Signal Processing, Ionospheric Signal Processing and Ionospheric Radio Propagation. She is head of the Commission H of URSI Turkey. She is also a member of the IONOLAB Group.

Faruk Erken received his undergraduate, graduate and Ph.D. from Firat University, Dicle University and Firat University in 1994, 2001 and 2014, respectively. He is currently an assistant professor at Department of Electrical and Electronics Engineering, Kastamonu University. His current research interests are Electrical Machines, Power Converters and Machine Learning. He is Representative Member of the Commission H of URSI Turkey.

Ali Cinar received his B.Sc. and M.Sc. in Electrical and Electronics Engineering from Atilim University and Eskisehir Osmangazi University in 2012 and 2019, respectively. He is currently working toward the Ph.D. in Signal Processing. He is currently research assistant with the Department of Electrical and Electronics Engineering, Kastamonu University.. His research interests include Digital Signal Processing and Deep Learning.