Classification of Ionospheric Scintillations during high Solar Activity and Geomagnetic Storm over Visakhapatnam Region using Machine Learning Approach

Shiva Kumar Nimmakayala* and Srilatha Indira Dutt V.B.S.*
Department of EECE, GITAM School of Technology, GITAM Deemed to be University, Visakhapatnam, INDIA
*121960402502@gitam.in; svemuri@gitam.edu

Abstract

The ionospheric plasma disturbances typically correlate with irregularities in electron density and ionospheric scintillations are produced in reaction to these variations generating radio signal fluctuations. Geolocation services and space based communication are endangered due to ionospheric scintillation which promptly produces fluctuations in information collected by Global Navigation Satellite Systems and this is at its strongest when the solar cycle is at its peak. Ionospheric space weather has a significant impact on Global Navigation Satellite Systems (GNSS) and one crucial aspect used in investigating ionospheric characteristics is total electron content (TEC). Due to fluctuations in time and space, the TEC obtained from GNSS signals is nonlinear and nonstationary.

In this study, machine learning approaches for Classification of the ionospheric scintillations were used during the high solar activity and geomagnetic storm in the month of July 2023. This approach enables the classification of ionospheric phase scintillations using well-known classifiers: Decision Tree and Support Vector Machine.

Keywords: Total Electron Content, Ionospheric Scintillations, Global Navigation Satellite System, Decision Tree, Support Vector Machine.

Introduction

Based on its altitude, chemical composition and temperature characteristics, the Earth's atmosphere is divided into layers. A layer of the atmosphere called the ionosphere is situated between 60 and 1000 kilometers above Earth's surface. It is a three-dimensional photo ionized dispersive media whose chemical composition fluctuates according to cosmic and solar radiation[14]. The day to day, seasonal, longitudinal and latitudinal variations, variations due to solar activity and variations due to magnetic activity all affect how charged particles are distributed in the ionosphere.

As a dispersive medium, the ionosphere prevents radio signals from traveling in a straight path[15]. In satellite-based positioning communication and navigation systems, the ionosphere may contribute to range rate errors which are a significant source of errors. The systems utilizing the ionosphere as a propagation medium and operating at specific frequencies like shortwave communications and microwave communications, are significantly impacted by the ionosphere's extremely fluctuating nature. The ionosphere's most significant parameter is the ionospheric TEC[5]. Along the signal transmission channel, it is determined as the integral of the electron density in a column with a 1 m^2 cross section. The total content electron (TEC) has a direct impact on how the ionospheric inaccuracy is corrected. It is an important source for research on ionospheric delay correlation and ionospheric estimation.

The inconsistencies in the electron density cause refraction and diffraction which cause the signal strength and phase to fluctuate whenever radio wave transmissions travel through unstable ionosphere and these fluctuations are referred as ionospheric scintillations. The solar and geomagnetic disturbance are the main causes of ionospheric irregularities[4,6]. When radio wave signals pass over anomalies in the electron density of the ionosphere, they experience rapid, random phase variations known as phase scintillations. The use of global navigation satellite systems (GNSS) is highly beneficial for understanding the earth's atmosphere and space weather. The GNSS services are susceptible to scintillation because it diminishes the services like precision availability and dependability. Therefore, developing classification of ionospheric scintillations is important. In order to classify the phase scintillations utilizing TEC, the support vector machine and decision tree classifiers, two of the most widely used supervised type learning techniques are employed in this study.

A machine learning technique called support vector method (SVM) is based on the idea of minimizing structural risk. One of the most used supervised type learning methods, the support vector machine method, is made to handle classification problems. The theoretically determined result will be the overall best option, guaranteeing strong generalization performance for unidentified samples and resolving the issue of local minimum for neural networks.

The neural network operates on statistics whenever the sample data goes to infinity, but the real sample data is restricted while the support vector machine is designed expressly for the circumstance of limited samples. The SVM algorithm cannot be used effectively with large data sets. SVM is ineffective when the data set has a significant quantity of noise such as overlapping target classes which

https://doi.org/10.25303/174da011017

occurs practically frequently[13]. When each data point's numeric feature value exceeds that of the training data samples, SVM underperforms. When there is a substantial amount of noise in the data set such as target classes that overlap often, SVM is unsuccessful. It is also ineffective for big data sets. There is no probabilistic reason for the classification that the SVM algorithm made.

Decision trees employ a non-parametric technique which means they make no assumptions about the geographical distribution. For the majority of classification problems, the decision tree (DT) algorithm which belongs to the supervised learning class of algorithms, is used. Using a non-parametric divide-and-conquer approach, the decision tree is another appreciated classifier that creates classification models using a tree topology[1]. It simultaneously divides the data set into smaller parts and builds a progressive decision tree. The structure of DT is systematically arranged with a set of rules being implemented in succession from the root node to the terminal node.

Choosing the relevance and significance of each of these features is crucial. In order to place the most pertinent feature according to the tree's rule, the data is separated at every node within the tree until a previously specified stopping point is satisfied. The decision tree is very easy to understand and does not require any sophisticated statistical expertise to be interpreted. It can also be employed during the data exploration stages because the decision tree algorithm is one of the fastest techniques for developing or detecting new features. It requires fewer procedures for data cleansing and unaffected by the data missing.

Data: The data was collected from the multi-constellation GNSS receiver at GITAM Deemed University Visakhapatnam (17.78160N, 83.37750E) and processed from 13th July 2023 to 19th July 2023 during the high solar activity and geomagnetic storm[7,8]. The elevation angle, TEC, time and Kp index are the variables considered in this investigation.

Material and Methods

Estimation of TEC: The TEC estimations are carried out using the following methodology in the research investigation of ionospheric scintillations. Each satellite from multi constellation broadcasts two carrier electromagnetic waves within the L1 (1575.42 MHz) and L2 (1227.60 MHz) frequency ranges. The difference between the L1 and L2 signal's ionospheric delays is measured by a dual-frequency GNSS receiver and the group delay can be expressed as follows:

$$P_2 - P_1 = 40.3 * TEC * \left[\frac{1}{f_2^2} - \frac{1}{f_1^2}\right] \tag{1}$$

where f_1 and f_2 are the appropriate high and low GNSS signal frequencies, the p1 and p2 denote the lengths of group path. The TEC from eq. (1) is written as[17]:

$$TEC = \frac{1}{40.3}\left[\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}\right](P_2 - P_1) \tag{2}$$

The 94 data samples collected every day at 15-second intervals are being used to estimate the following TEC value. The phase scintillation index is subsequently determined by utilizing the estimated TEC.

Phase Scintillations: There are two sorts of phase scintillation computations.

1. Computation
2. Classification

The phase scintillation index is subsequently calculated using the estimated TEC from eq. (2). The GNSS signal's phase fluctuation induced by the transit of the ionosphere contains the following refractive component [9]:

$$\emptyset = \frac{Q^2}{2C\varepsilon_0 M_e F(2\pi)^2} \int N_e d_\rho \tag{3}$$

where "C" signifies the light speed, "$\int N_e d_\rho$" signifies the total electron content, "Q" signifies the electron's charge, "F" signifies the frequency expressed in hertz, "M_e" signifies the electron's mass and "ε_0" signifies the time-dependent free space permittivity.

The following reduction of eq. (3) is achieved utilizing MKS units:

$$\emptyset = \frac{40.3}{CF} TEC \tag{4}$$

where "C" signifies the light speed, "TEC" signifies the total electron content and "F" signifies the frequency expressed in Hertz.

The phase scintillation index ("σ_Φ") is characterized as the standard deviation of "$\emptyset$" in radians which is as follows. This index is used for accumulating phase variation observations.

$$\sigma_\Phi = \sqrt{\frac{\sum(\emptyset_i - \emptyset\mu)^2}{N}} \tag{5}$$

The efficiency of its index serves to define the characteristics of the phase scintillations. The weak phase scintillations occur when the "σ_Φ" is less than 0.3, moderate phase scintillations occur when the "σ_Φ" is between 0.6 and 0.3 and strong phase scintillations occur when the "σ_Φ" is greater than 0.7.

Classification: The following are classifications of the phase scintillations made using the support vector method and decision tree.

Support Vector Machine (SVM): The support vector machines (SVM) in machine learning are the supervised learning models with corresponding learning strategies.

https://doi.org/10.25303/174da011017

SVMs, which rely on statistical learning architectures, are among the most reliable classification techniques. A collection of training examples is provided, each of which is flagged as falling into one of two categories and in order to categorize fresh instances, new samples are mapped into the identical space and classified to belong to a category[3].

$$F(X_{SVM}) = y_{SVM} = W^T X_{SVM} + b \quad (6)$$

where b signifies the bias, $w = [w_1, w_2, ..., w_M]$ signifies the weight vector and X_{SVM} signifies the input vector.

Decision Tree (DT): A structure known as a decision tree includes two or more child nodes for each internal node, going all the way up to leaf nodes which are nodes without any children. Every internal node corresponds to a test of an attribute and the child branches which emanated from that node show the potential outcomes of that test.

Every single leaf node located at the tree's tip includes a classification identifier attached to it. Pruning strives for and eliminates any tree branches that are unnecessary or repetitive for classifying the outcome [10].

$$\Delta i(s,t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (7)$$

where "s" signifies a potential split at each of the nodes t that splits both a left (t_L) and right (t_R) child nodes in relation to p_L as well as P_R. In this instance, the splitting of the impurity standard i(t) is specified. The most accurate measure of impurity removal form split "s" is known as the $\Delta i(s,t)$.The three impurity measurements that are typically employed are Gini index, Chi-square and Gain ratio. Between 0 to 1, the Gini impurity index (Ig) is available.

$$I_g\,(t_{U(u_i)}) = 1 - \sum_{k=1}^{m} f(t_{U(u_i)}, k)^2 \quad (8)$$

where $f(t_{U(u_i)}, k)$ is the u_i probability for each sample departing at node t from sample k. The criteria for the decision tree splitting are based on the Ig value with the lowest value.

Performance metrics for Classification: The current research evaluates the performance of these machine-learning-based classification models using a variety of performance indicators. It should be highlighted that the effectiveness of these classification models cannot be determined by a single metric. Since accuracy, precision, recall and F1 score are the performance indicators, a comprehensive evaluation system should also include them[2].

Accuracy: The model's functionality is evaluated using accuracy. It computes the proportion of precise events relative to all possibilities.

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (9)$$

Precision: The precision of a model determines how precise its optimistic estimates are. This statistic is the ratio of accurate positive estimates to all favorable estimates produced by the model.

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (10)$$

Recall: Recall gauges a classification model's ability to accurately extract all relevant instances from a dataset. It measures the proportion of true positives (TP) to true positives as well as false negative (FN) occurrences overall.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (11)$$

F1 score: A classification model's overall efficacy is evaluated using the F1score. The average of both recall and accuracy is used to calculate it.

$$\text{F1-score} = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (12)$$

Figure 1 depicts the methodology flow chart for the research. Data is gathered and analyzed during the high solar activity and geomagnetic storm in the month of July 2023 from the multi-constellation GNSS receiver at GITAM Deemed to University Visakhapatnam (17.78160N, 83.37750E). The TEC is calculated during the same duration using this GNSS data and using an estimated TEC, the phase scintillations are computed. The phase scintillations are classified by using the widely utilized algorithms: SVM and decision tree. Using the classifiers' performance metrics, the results are compared.

Results and Discussion

In order to calculate and analyze TEC estimations and classify phase scintillations, the multi-constellation GNSS receiver's carrier phase measurements data are taken into account and the data are sampled at 15 seconds. In the present research, range measurements are used to determine TEC. Phase scintillations are then calculated using the estimated TEC. Using decision trees and SVM, the phase scintillations are finally classified. This procedure is carried out because the process is performed during the high solar activity and geomagnetic storm in the month of July 2023.

TEC Estimation: In the present research, range measurements are used to estimate TEC using eq. (2). TEC values for one week are estimated using one-day TEC data. This procedure is carried out for the process performed during the high solar activity and geomagnetic storm in the month of July 2023. The total electron content for a week during the high solar activity and geomagnetic storm in the month of July 2023 is shown in figure 2. With the help of range measurements made during the high solar activity and geomagnetic storm in the month of July 2023 and data from one day's TEC, figure 3 shows the estimated TEC for one week.

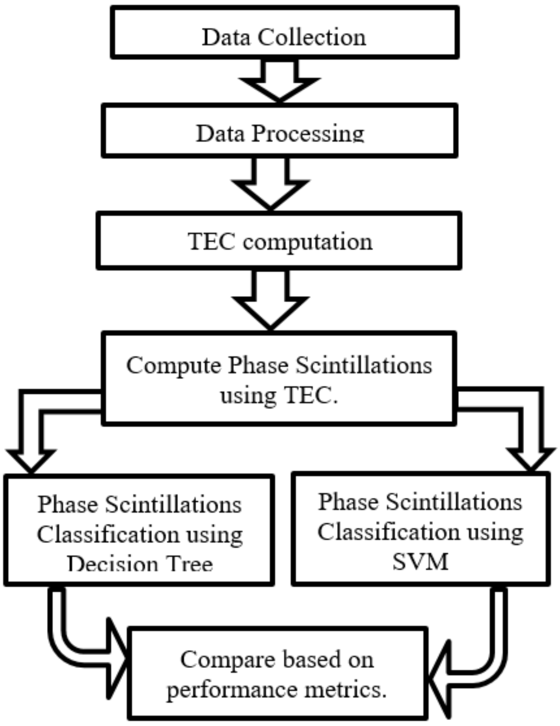

Figure 1: Flowchart for Methodology

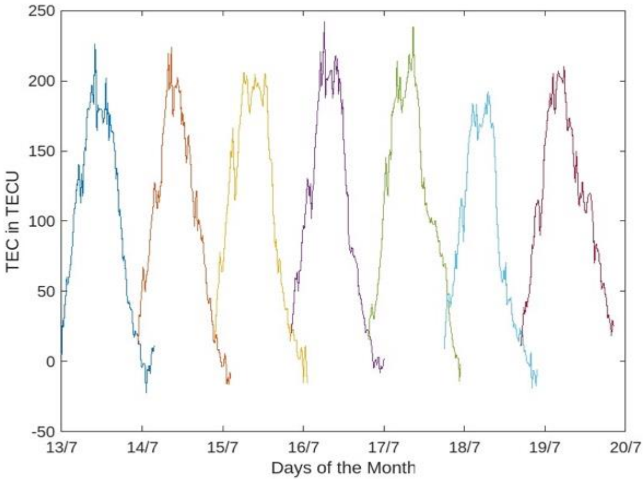

Figure 2: TEC for a week

Phase Scintillations Computations: With the help of TEC estimates and range measurements, the phase scintillations are calculated. The changes in phase scintillations during the high solar activity and geomagnetic storm in the month of July 2023 are depicted in figure 4 and they may be classified into three levels: low, moderate and strong using widely recognized machine learning classifiers namely Decision Tree and Support Vector Methods (SVM). The range of the 10.7 radio flux and the sunspot number for July 2023 is 141.8 to 149.8 and 110.4 to 120.4 respectively[11]. High scintillations in the month of July 2023 are anticipated because of the high sunspot number and 10.7 radio flux[13]. High scintillations are anticipated in the month of July 2023 as Solar Cycle 25 approaches its maximum[12].

Phase Scintillations Classifications: The decision tree and SVM and are used to classify the phase scintillations. Figure 5 illustrates a decision tree classifier for phase scintillations classification during the high solar activity and geomagnetic storm in the month of July 2023 with 161 samples as the sample size (data as of July 2023). The confusion matrix depicted in figure 6 is used to assess how well the decision tree performed during the high solar activity and geomagnetic storm in the month of July 2023.

The decision tree's confusion matrix for the phase scintillations of a 1-week data during the high solar activity and geomagnetic storm in the month of July 2023 is shown in fig. 6. The week dataset consists of 161 samples, all the samples are correctly identified.

The confusion matrix depicted in fig. 7 is used to assess how well the SVM performed during the high solar activity and geomagnetic storm days in the month of July 2023.

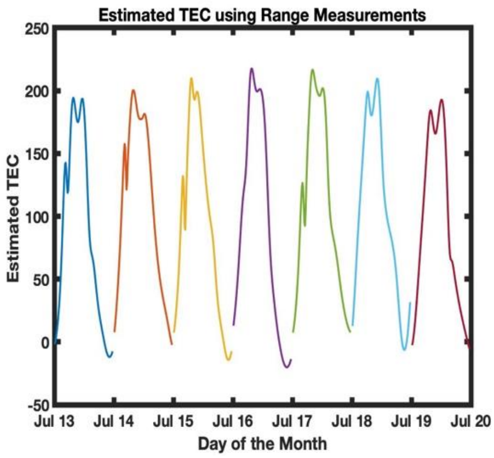

Figure 3: TEC Estimated using Range measurements

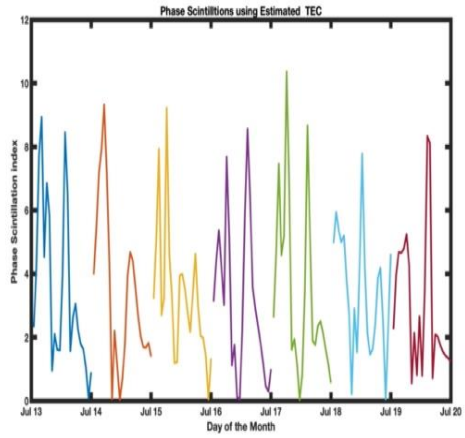

Figure 4: Phase Scintillations for Estimated TEC

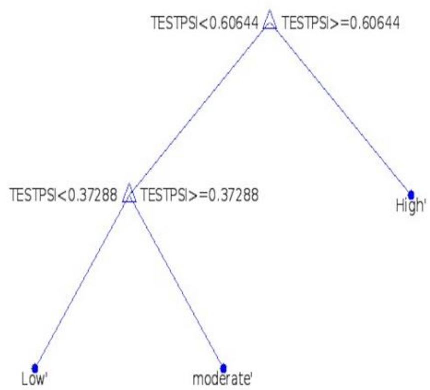

Figure 5: Classification of Phase Scintillations during a Decision Tree Classifier

Table 1
Decision Tree and SVM Performance Metrics

Model	Sample Size	Accuracy	precision	Recall	F1 score
SVM	161 (1 week)	98.13%	100%	98.01	98.99
DT	161 (1 week)	100%	100%	100%	100%

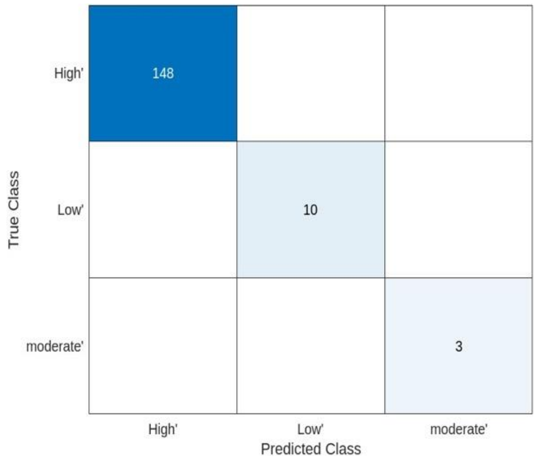

Figure 6: Confusion matrix for 1 week using Decision Tree

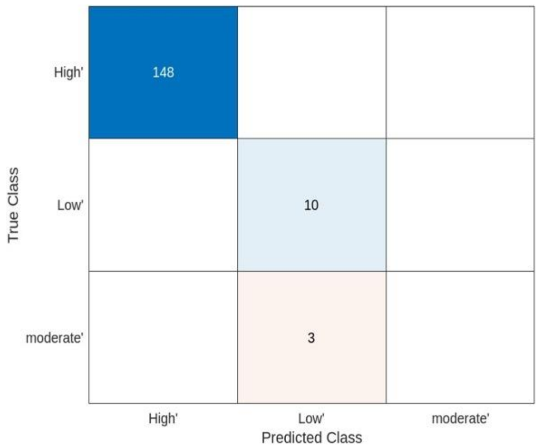

Figure 7: Confusion matrix for 1 week using SVM

The SVM's confusion matrix for the phase scintillations of a 1-week data during geomagnetic storm the high solar activity and geomagnetic storm days in the month of July 2023 is shown in figure 7. The week dataset consists of 161 samples, 158 samples are correctly identified. The performance of SVM and Decision Tree classifiers is compared using the performance metrics: accuracy, precision, recall and F1 score. The comparison is displayed in table1.

Conclusion

In the present research, we investigate the feasibility of estimating total electron content (TEC) via data assimilation and ionospheric scintillation Classification is provided. Using range measurements, the TEC over the Visakhapatnam region is estimated. The computation of ionospheric scintillations is carried on by using estimated TEC. Decision Tree and SVM classifiers are used to categorize ionospheric scintillations. Since the TEC variation is dependent on the solar activity and geomagnetic storm, the ionospheric scintillations are more prominent during the storm.

The decision trees and SVM are used to classify ionospheric scintillations and confusion matrices are used to evaluate them. The accuracy of the decision tree is 100% during a geomagnetic storm, compared to the accuracy of the SVM, which is 98.13%. This shows that the decision tree works better than the SVM for classifying ionospheric scintillations. Therefore, it can be inferred that the decision tree is strongly preferable to the widely recognized SVM classifier for classifying ionospheric scintillations. Consequently, the classification of ionospheric fluctuations and scintillations can be thought of as a potential application of machine learning techniques.

References

1. Cravioto D.A.G. et al, Analysing Factors That Influence Alumni Graduate Studies Attainment with Decision Trees, 2020 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Iraq, 44-49 **(2020)**

2. Dey A., Rahman M., Ratnam D.V. and Sharma N., Automatic Detection of GNSS Ionospheric Scintillation Based on Extreme Gradient Boosting Technique, *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5 **(2022)**

3. Guozhen Xia et al, Ionospheric TEC forecast model based on support vector machine with GPU acceleration in the China region, *Advances in Space Research*, **68**, 1377-1389 **(2021)**

4. Guozhu Li et al, Effects of geomagnetic storm on GPS ionospheric scintillations at Sanya, *Journal of Atmospheric and Solar-Terrestrial Physics*, **70**, 1034-1045 **(2008)**

5. Hussain F. et al, Complexity and Limitations of GNSS Signal Reception in Highly Obstructed Environments, *Eng. Technol. Appl. Sci. Res*, **11** ,6864–6868 **(2021)**

6. Han Y. et al, Machine Learning-Based Short-Term GPS TEC Forecasting During High Solar Activity and Magnetic Storm Periods, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **15**, 115-126 **(2022)**

7. Hindustan times, https://tech.hindustantimes.com/tech/news/beware-solar-storm-to-strike-earth-today-noaa-warns-716893198 13220.html, accessed on 28-08-2023 **(2023)**

8. Kompas, https://www.kompas.id/baca/english/2023/07/18/en-badai-geomagnetik-menghantam-bumi-pada-selasa-18-juli-2023, accessed on 28-08-2023 **(2023)**

9. Kintner P.M. and Ledvina B.M., GPS and ionospheric scintillations, Space Weather **(2007)**

10. Linty N. et al, Detection of GNSS Ionospheric Scintillations Based on Machine Learning Decision Tree, *IEEE Transactions on Aerospace and Electronic Systems*, **55**, 303-317 **(2019)**

11. NOAA, https://www.swpc.noaa.gov/products/predicted-sunspot-number-and-radio-flux, accessed on 28-08-2023 **(2023)**

12. NOAA, https://www.swpc.noaa.gov/products/solar-cycle-progressionK, accessed on 28-08-2023 **(2023)**

13. Paul B.D.E.A. and Guha A., Comments on the percentage of occurrence methodology used in "a study of L band scintillations during the initial phase of rising solar activity at an Indian low latitude station" by H J Tanna, S P Karia and K N Pathak, *Advances in Space Research*, **63**, 1227-1233 **(2019)**

14. Shiva Kumar N. and Vemuri B.S.S.I.D., Analysis of Ionospheric Scintillations using GPS and NavIC Combined Constellation, *Eng. Technol. Appl. Sci. Res*, **13**, 10936–10940 **(2023)**

15. Sirish Kumar P. and Srilatha Indira Dutt V., Development and performance evaluation of Correntropy Kalman Filter for improved accuracy of GPS position estimation, *International Journal of Intelligent Networks*, **3**, 1-8 **(2022)**

16. Seyyed Reza Ghaffari-Razin et al, Modeling and forecasting of ionosphere TEC using least squares SVM in central Europe, *Advances in Space Research*, **70**, 2035-2046 **(2022)**

17. Ya'acob et al, Total Electron Content (TEC) and Estimation of Positioning Error Using Malaysia Data, Lecture Notes in Engineering and Computer Science, Proceedings of the World Congress on Engineering **(2010)**.

(Received 29th November 2023, accepted 09th January 2024)