Research Article

Deep learning of total electron content

Omid Memarian Sorkhabi[1]

Abstract

One of the most notable errors in the global navigation satellite system (GNSS) is the ionospheric delay due to the total electron content (TEC). TEC is the number of electrons in the ionosphere in the signal path from the satellite to the receiver, which fluctuates with time and location. This error is one of the major problems in single-frequency (SF) GPS receivers. One way to eliminate this error is to use dual-frequency. Users of SF receivers should either use estimation models or local models to reduce this error. In this study, deep learning of artificial neural networks (ANN) was used to estimate TEC for SF users. For this purpose, the ionosphere as a single-layer model (assuming that all free electrons in the ionosphere are in this thin layer) is locally modeled by the code observation method. Linear combination has been used by selecting 24 permanent GNSS stations in the northwest of Iran. TEC was modeled independently of the geometry between the satellite and the receiver, called L4. This modeling was used to train the error ANN with two 5-day periods of high and low solar and geomagnetic activity range with a hyperbolic tangential sigmoid activation function. The results show that the proposed method is capable of eliminating ionosphere error with an average accuracy of 90%. The international reference ionosphere 2016 (IRI2016) is used for the verification, which has a 96% significance correlation with estimated TEC.

Keywords GNSS · Total electron content · Deep learning · Artificial neural networks

1 Introduction

The ionosphere has extended from an altitude of 80 km to more than 1000 km. This layer of the atmosphere has important and fundamental effects on radio waves and their transmission. Radio waves travel slower when passing through this layer, resulting in longer transmission time. This is called the delay in the transmission of waves, which must be precisely determined for the accurate positioning done [1]. The ionosphere is a dispersive medium, and the dispersion of the waves depends on their frequency. Dual-frequency global positioning system (GPS) receivers reduce the effect of this error by accurately positioning [2]. But these receivers are expensive and partly complex,

most users and systems use SF GPS receivers, which are not capable of eliminating or reducing the effect of the ionospheric delay [3]. For this purpose, it will be important for ionospheric models to be able to accurately represent ionosphere characteristics to allow for accurate positioning using GNSS measurements [4]. The signal transmission delay is proportional to the free electron distribution in the signal transmission path from the satellite to the receiver. Generating TEC maps directly from GPS measurements is one of the methods for studying the ionosphere [5].

Numerous studies have been conducted to investigate ionosphere error using different empirical models [6, 7]. The capability of ANN in this area is evaluated by temporal TEC estimation and presented as a suitable modeling

✉ Omid Memarian Sorkhabi, omidmemaryan@gmail.com | [1]Department of Geomatics Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, 81746-73441 Isfahan, Iran.

tool [8]. TEC IRI2012 estimate of GPS models in the low-latitude East Africa regions has good consistency and accuracy of over 75% offers [9]. The estimated TEC from GPS based on IRI2016 long short-term memory DL models is $R2 = 0.99$ [10]. The estimated GNSS TEC with ionosonde outputs in various distributions is highly consistent, with the most changes estimated in the 9 March 2012 magnetic storm [11].

Radial basis function ANN is a reliable tool for predicting TEC and its average relative error is 9% [1]. Regulated DL can estimate TEC at different locations, times, and different geomagnetic conditions of the ionospheric structure and is effective in filling the missing data [2]. The TEC profile estimated with the ANN in 3D is in good agreement with the Jicamarca and Millstone Hill stations radar sensors. The proposed model represents the daily, seasonal, and annual variation of the ionosphere successfully [13].

The innovation of this paper is the use of artificial intelligence with special architecture in order to estimate TEC with high accuracy. In this study, DL was used to estimate TEC for SF users. First, the mathematical equations of local ionosphere computation and the propagation error of ANN is presented. Estimated TEC from GNSS with 24 stations in northwest Iran has been investigated using code observations in two high and low SGA time zones. The numerical results are then compared with the TEC values of IRI2016.

2 VTEC (TEC)

The first-order ionosphere delay is a function of the integral expression $\int N \cdot ds$, which is called TEC. TEC is the number of electrons in the ionosphere in the signal path from the satellite to the receiver, expressed in TECU units (each electron per square meter $TEC = 10^{16}$). Using TEC, we can write the first-order delay according to relation 1 [13].

$$I_{j,g}^{(1)} = \frac{A}{2f_j^2} TEC \tag{1}$$

where $A = 80.6$ m³/s² and f is the carrier signal frequency. TEC has very high variations due to changes in time and space. Another important point is that the TEC is strongly dependent on the satellite's elevation angle (the geometric position of the satellite) because the signal path length in the ionosphere varies with the satellite position in the sky [14].

2.1 Modeling the ionosphere locally

Ionosphere modeling involves steps such as collecting ionosphere measurements, data processing, analysis, and finally validation and verification of results. Different models can be used to model the ionosphere. The model used in this study is local for modeling the ionosphere in Northwest Iran. The mechanism of this model is that it considers all free electrons in a thin layer about 450 km from the Earth's surface [10]. This model is known as the single-layer model. The mapping function of this model is written F_I with relationships 2 and 3 [9].

$$F_I(Z) = \frac{E}{E_V} = \frac{1}{\cos Z'} \tag{2}$$

$$Sin Z' = \frac{R}{R + H} Sin z$$

where Z' and Z are the zenith distance at single-layer and station height, respectively. R is the mean radius of the Earth and H is the height of the single layer above the Earth surface. At the height of this ideal layer, it is expected to have the highest electron density. In addition, the electron density E at this surface is a function of the latitude or geomagnetic latitude and the constant solar length S [7]. To produce the TEC map, geometry-free linear components called $L4$, which contain ionospheric information, are used. The nondifferential observational equations of the phase and code are written with relationships 4, 5, and 6 for the calculation of the vertical local TEC [11].

$$L4 = -a\left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right)F_I(Z)E(\beta, s) + B_4 \tag{4}$$

$$P4 = +a\left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right)F_I(Z)E(\beta, s) + b_4 \tag{5}$$

$$E_V(\beta, s) = \sum_{n=0}^{n_{max}} \sum_{m=0}^{m_{max}} E_{nm}(\beta - \beta_0)^n (s - s_0)^m \tag{6}$$

where $P4$ and $L4$ are phase and code observations free of geometry. $a = 40.3 \times 10^{17}$ ms⁻² TECU⁻¹ is constant. $F1$ and $F2$ are the frequencies of the wave phases carrying $L2$ and $L1$, respectively. $F_I(Z)$ is a function of the mapping evaluated at zenith distance Z' [1, 12]. $E_V(\beta, s)$ TEC is a function of s (latitude) and β (longitude). s_0 and β_0 are the reference latitude and reference longitude. a constant bias (in meters) is B1 due to the ambiguity of the inner phase ambiguity and B2 to the wavelengths $\lambda 1$ and $\lambda 2$ [6]. The

reason for using pseudo observations is low noise and multi-path error compared to phase equations.

2.2 Kp-index

Geomagnetic activity in the field of ionospheric research is measured by the Kp-index, which indicates solar radiation. The Kp-index is obtained by averaging the observations of the horizontal field H intensity from a network [3]. This index is calculated at 3-h intervals, which is a kind of global measurement of the magnetic deviation from regular daily changes in one-hour period [14].

3 Deep learning of artificial neural networks

Artificial Neural Networks with simple processor units are capable of parallel processing, storing knowledge and using it for sequential evaluation [15]. These networks are a simplified model of human brain decision making, formed by simple, synthetic neurons [16]. The input information of the neural cells is managed by the mean synaptic weight, which is known as the learning process during the sequential bursting process. After the training process with the activation function, this process is used throughout the neural cell to produce information [1].

ANN with a hidden layer and sigmoid activation functions in the middle layer and linear conversion functions in the output layer will be able to approximate the desired functions with any degree of approximation, provided they are sufficient. The middle hidden layer has several neural cell. The activation function used in the hidden neural cell is the sigmoid function of the hyperbolic tangent, expressed as a relation 7 [17].

$$f(z) = \frac{2}{1 + e^{-2z}} - 1 \tag{7}$$

where z is the input information of the neural cell and $f(z) \in [0, 1]$. The output and input values of the neural networks are defined in this range. This function must be continuous, derivative and monotonically descending [16]. It is also expected to have a saturate activation function (asymptotically approaching its maximum and minimum values).

A backpropagation artificial neural network (BPANN) error propagation algorithm is proposed for estimating the TEC because it has a good estimation function and has some estimator quality criteria such as inaccuracy, compatibility, minimum variance, efficiency and adequacy [18, 19].

3.1 BPANN

This method is more widely used in engineering sciences than other methods of artificial neural networks. BPANN is a supervised learning network and feed forward. The network with a hidden layer and the use of the sigmoid activation function can estimate any continuous function by taking a sufficient number of hidden neurons [20]. The BPANN training process, similar to adjustment, attempts to reduce the residual output of the network [21]. This is done by initializing the weights and communicating between the neural cells of each layer. To prevent BPANN learning from slowing down, the initialization of weights between 0 and 1 is usually selected.

The Delta rule is used based on the least-squares error in BPANN. Training process with a set of data consisting of specified input and output parameters based on hidden layer and weighting layer adjustment [22]. This iteration works by updating the weights and reducing the residuals of the neural network output (the difference between the calculated output and the actual output) and has two main feedback and propagation steps. These steps are performed continuously with a set of training data of over several thousand iterations. In the feed-forward step, each input unit receives an input signal and sends it to each of the hidden units. Each hidden unit then calculates its activation and sends its signal to all output units [23]. A supervised value is also available in the tutorial for each input pattern. During the backpropagation training, each activation output unit compares its computational activation output with its target value to determine its associated error and to reduce the error value in the backpropagation step [23, 24]. Mean square error (MSE) can be used as a measure of neural network efficiency. For a set with N inputs, MSE is defined as a relation 8.

$$MSE = \frac{\sum_{i=1}^{N} \left(y_i^{act} - y_i^{pred} \right)^2}{N^2} \tag{8}$$

where y_i^{act} represents the actual output and y_i^{pred} represents the output estimated by the neural network. The BPANN network architecture used in this study is illustrated in Fig. 1, which consists of three input layers of latitude, longitude and, time, hidden layers (with hyperbolic tangential sigmoid activation function) and a vertical TEC output layer. Seventy percent of the data were used for training, 15% for testing, and 15% for validation. Further details on the BPANN learning process can be found in Bishop [20], Yang et al. [25] and Gupta and Singh [22].
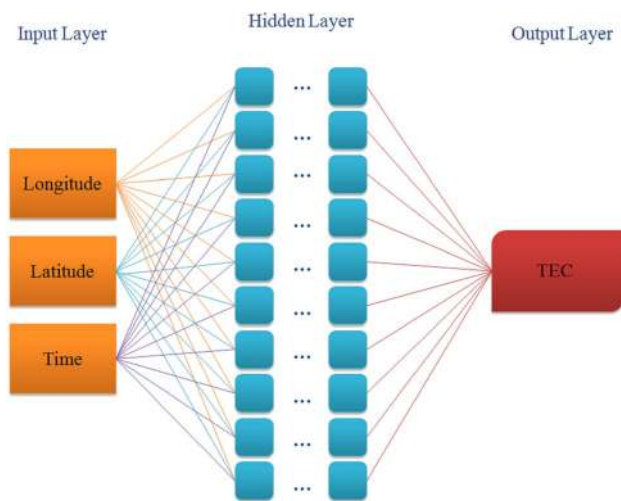
Fig. 1 BPANN architecture

4 The study area and results

Geodynamic GNSS stations of Iran with 127 stations have been designed and implemented by the national cartographic center since 1994 to monitor the earth's crust movements. The study area is northwest of Iran with latitude and longitude $37 \leq \varphi \leq 40$ and $45 \leq \lambda \leq 48$ and an approximate area of 150,000 square kilometers. The study used 24 GNSS stations of the East Azerbaijan region that collect 24-h data. GNSS data were processed using the standard BERNESE software process. The processing uses the IGS (International GNSS Service) final satellite orbit data and the Earth's rotational parameters. To model the ionosphere and estimate the TEC, the ionosphere is considered as a thin layer around the Earth with a constant height from the earth, assuming that all free electrons in the ionosphere are contained in this thin layer. Figure 2 shows the distribution of stations used in this study. All GNSS permanent stations have recorded observations with Ashtech Z12 and Trimble 4000SSI receivers with the choke-ring antenna.

Two data series with $Kp \geq 4$ and $Kp \leq 4$ SGA were used to estimate the DLANN. For the period of high geomagnetic and solar activity (PHGSA) ($Kp \geq 4$) the ranges of 13 July 2012 to 17 July 2012 and for the period of low geomagnetic and solar activity (PLGSA) ($Kp \leq 4$) ranges 8 August 2012 to 12 August 2012 have been considered. Figure 3a shows the geomagnetic activity from 11 July 2012 to 15 August 2012. The Kp-index had the highest value on 15 Jul 2012 and the lowest value on 10 July 2012, which is
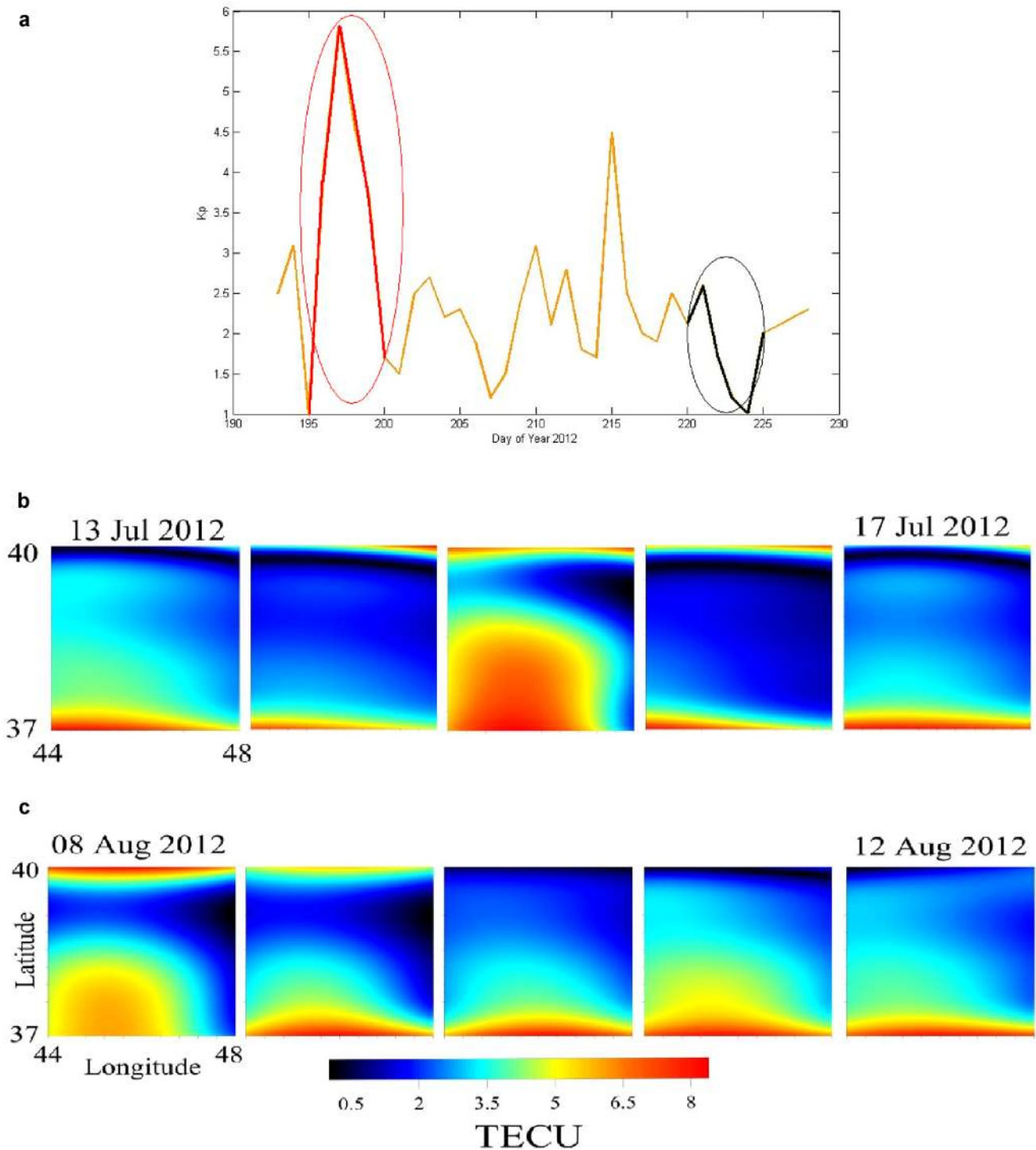
Fig. 2 Distribution of GNSS permanent stations
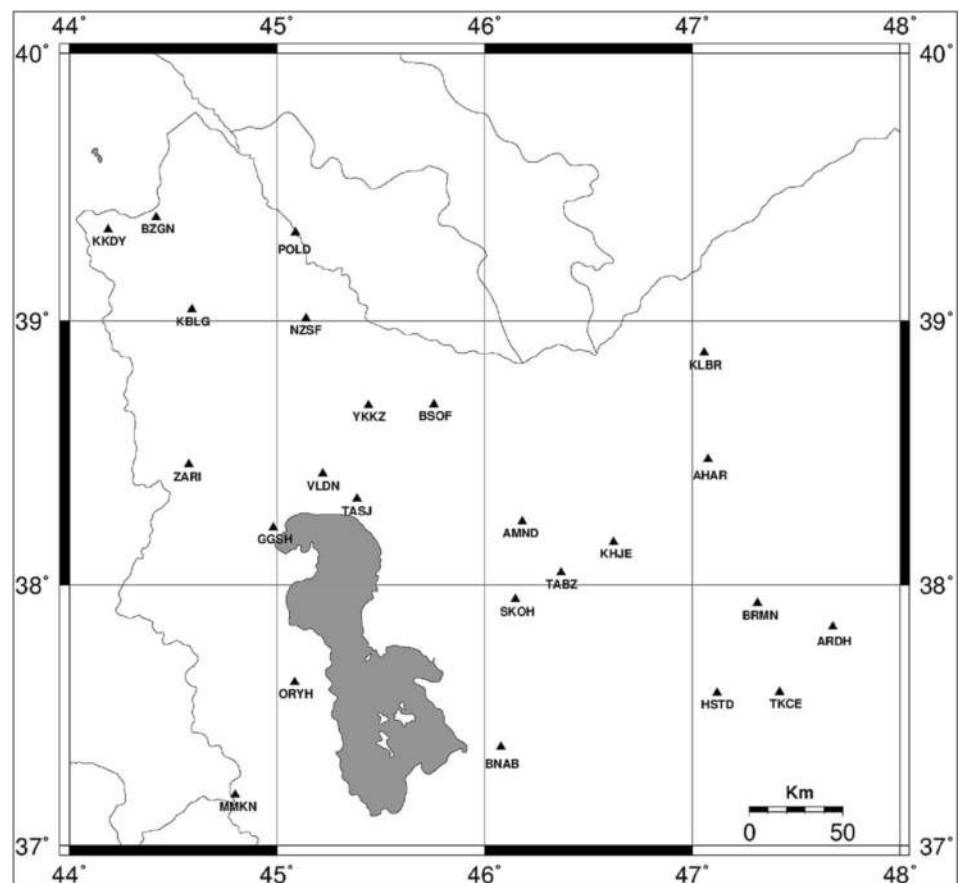
Fig. 3 a Geomagnetic activity from 11 JUL 2012 to 15 AUG 2012, red and black ellipse shows the range of PHGSA and PLGSA processed in this study, respectively **b** TEC of PHGSA **c** TEC of PLGSA

also evident in the estimated TEC. The range of TEC variations in the study ranges from 0 to 8 TECU at low latitudes greater than high latitudes. According to Fig. 3b, c, TEC has the highest value on 15 July 2012 and 8 August 2012. At lower latitudes, the TEC is higher than the higher latitudes.

Also in the northeast of the study area in July and August, TEC is lower than in other areas.

BPANN trains the network with the gradient method to minimize MSE. Increased repetition in network training increases network accuracy. But too much repetition will not

have much effect on increasing network accuracy. Increasing the number of neural cells in the middle layer increases the accuracy of the neural network, but excessively increasing the number of neural cells decreases the accuracy of the network because any increase will not always improve the network [26]. MATLAB has been used to implement DL. According to the hyper-parameters, hidden layers with changes in different training have been optimally selected. Also, the data for the training, testing and validation steps were randomly selected at different times.

To estimate the accuracy of backpropagation error, the absolute and relative errors in the form of nine and ten relationships are used.

$$\text{Absolute Error} = |TEC_e - TEC| \qquad (9)$$

$$\text{Relative Error} = \frac{\text{Absolute Error}}{TEC} \times 100 \qquad (10)$$

where TEC_e is the estimated value (the value obtained after the BPANN) and the TEC is the calculated value (the value obtained from the L4 code equations). In this study, 428 estimates were made for the TEC north-west of Iran between different stations, days and hours. The mean absolute error of the estimates is 1.4 TECU with a standard deviation of 1.1 TECU and the mean relative error of 11.8% with a standard deviation of 10.3%. Figure 4a, b shows the absolute error and the relative error during PLGSA, respectively. Figure 4c, d shows the absolute error and the relative error during PHGSA, respectively. In PLGSA, the absolute error is greater in the northwest and southeast regions, and the relative error is greater in the western and southeast regions. In PHGSA, the absolute error is greater in the northwest and southeast regions and the relative error in the northern, eastern, western and southern regions is higher. In the central regions of both solar activity periods, absolute error rates are observed due to the high density of the GNSS stations, which is similar to the relative error. ANN can estimate
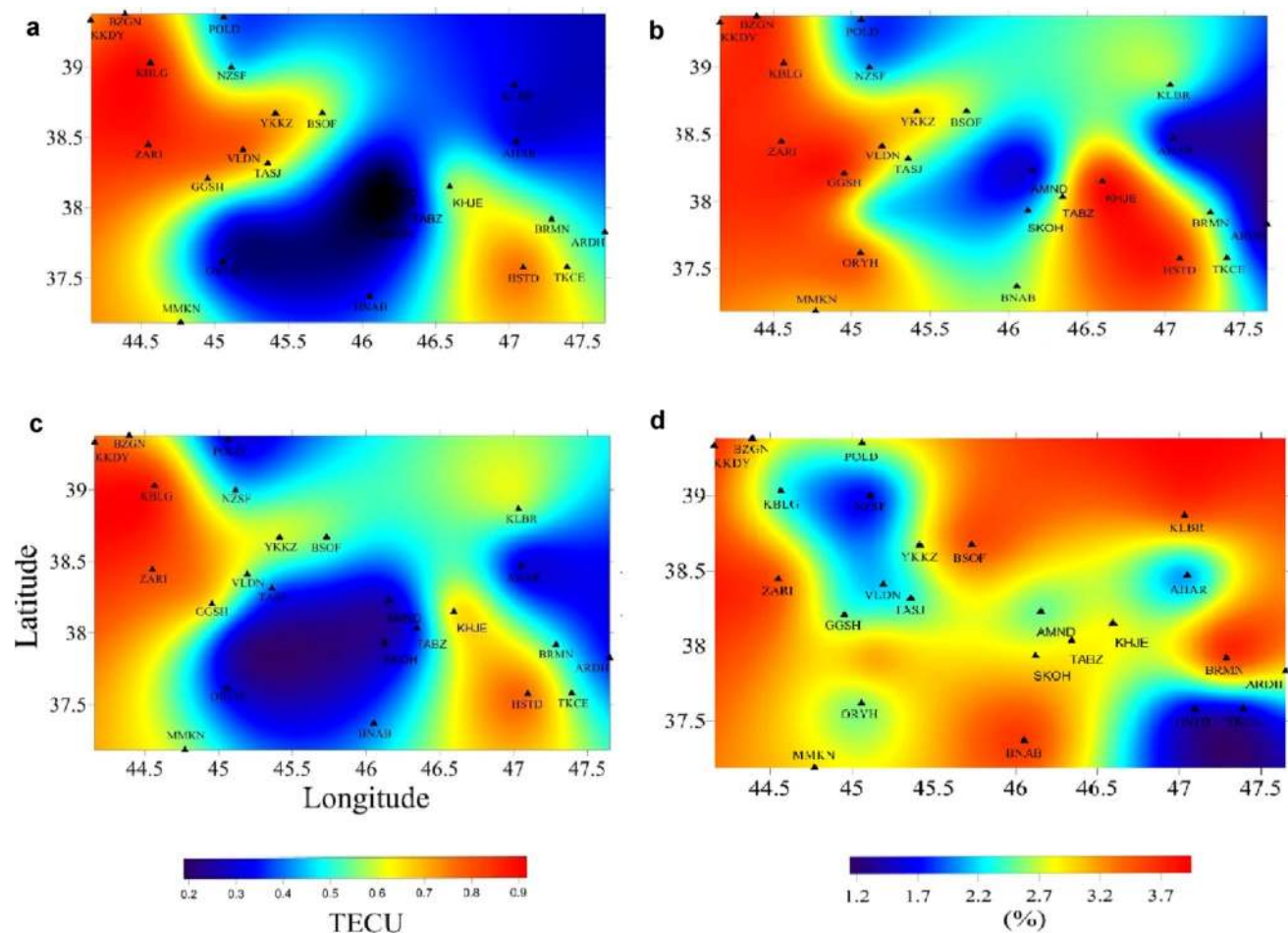

Fig. 4 **a** Absolute error during PLGSA **b** relative error during PLGSA **c** absolute error during PHGSA **d** relative error during PHGSA

the required parameters with high accuracy due to the simultaneous usability of several factors, they also do not require complex mathematical formulas and can be properly selected and trained if the network pattern is correct (good for estimating parameters in a short time). Based on the accuracy, BPANN is a very powerful network that can be used to estimate various parameters. These networks will be able to provide logical answers to new input data if they are properly trained [2, 5, 14].

To evaluate the accuracy of the proposed DL model, a support vector machine (SVM) with different kernels has been used to identify features (longitude, latitude and time) and forecast, the results of which are given in Table 1. The fine Gaussian kernel is more accurate than the rest and the lowest is the linear kernel. Due to the accuracy of about 90%, DL has a successful performance compared to the SVM method and to improve the accuracy of SVM, other features must be defined. Figure 5a shows the accuracy of different methods.

The IRI empirical model is an international project and the result of the collaboration of the Committee On Space Research (COSPAR) and the International Union of

Radio Science (URSI), which set up a working group with the aim of establishing an international standard for the determination of ionospheric parameters as the IRI model. The experimental IRI model was developed as a numerical model to avoid the complexities of theoretical models based on all available data sources [27]. For validation, the IRI2016 model was used in the study period. The IRI-2016 model is suitable for scientific analysis of the overall behavior of the ionosphere [28]. The results show an almost linear relationship between TEC from IRI 2016 and GNSS and their SC is 96% and their difference is less than 2 TECU. An almost linear relationship is observed between IRI 2016 and TEC estimated from GNSS and in PHSA variations with a value greater than 8, it also provides consistent results. Also at lower TEC values is good compatibility between IRI 2016 and TEC estimated from GNSS.

5 Conclusion

Two data series with Kp ≥ 4 and Kp ≤ 4 SGA over 5 days were used to estimate the TEC. The results show that the mean absolute error of the estimates is 1.4 TECU with a standard deviation of 1.1 TECU and the average relative error of 11.8% with a standard deviation of 10.3%. In this study, the ability of DLANN has estimated the TEC that is based on BPANN is close to the real values. It is also capable of eliminating ionosphere error with an average accuracy of 90%. Estimated values can be used by users of SF receivers in the study area to correct ionosphere error. With the increase in the number of hidden layer neurons and passing the current set point, the network with the overtrain phenomenon and the MSE value will increase

Table 1 Accuracy of different SVM kernel functions

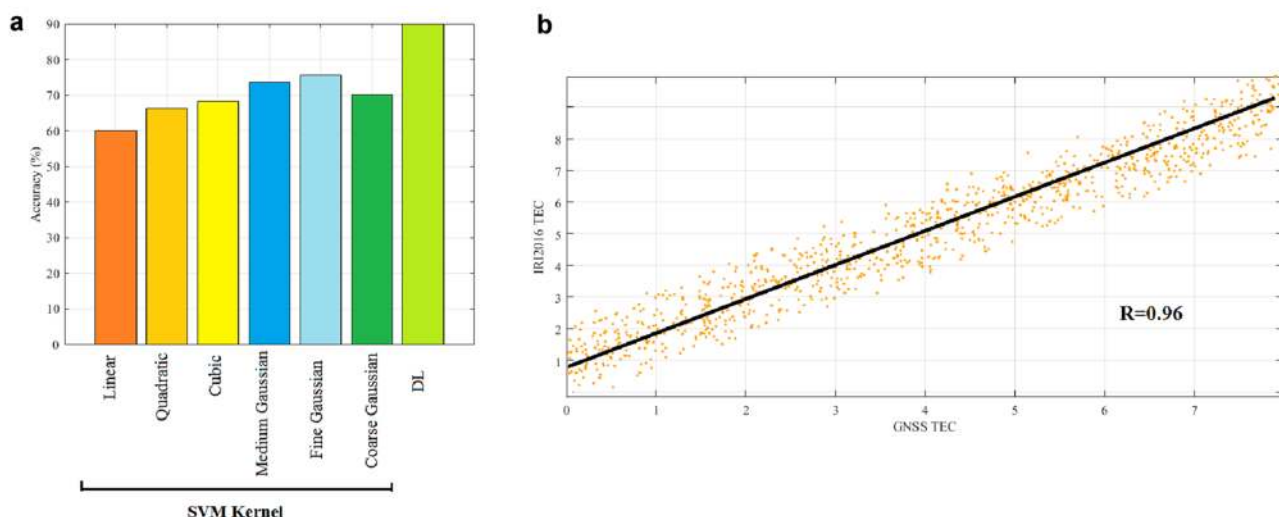| SVM Kernel	Accuracy (%)
Linear	60.12
Quadratic	66.41
Cubic	68.24
Medium Gaussian	73.74
Fine Gaussian	75.68
Coarse Gaussian	70.14

Fig. 5 a Accuracy of different methods **b** scatter plot of TEC from IRI2016 and GNSS during the study period

sharply. Therefore, the best number of neural cells or attention to minimizing MSE for network training should be calculated using the trial and error process. The number of hidden layers required is determined by the convergence of the training process. The disadvantages of DL are the inability to interpret the outputs and how to select training data, which can be used to solve this problem with the genetic algorithm (find the optimal number of training data). Further studies can estimate TEC values in other regions using different activation functions and compare the efficiency and accuracy of DL with local and global ionosphere models. The proposed method can be useful in other fields of geodesy. Therefore, it is recommended to implement in other areas such as geoid determination and gravity and compare the results with wavelet [12].

Declarations

Conflict of interest The author declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Huang Z, Yuan H (2014) Ionospheric single-station TEC short-term forecast using RBF neural network. Radio Sci 49(4):283–292
2. Chen Z, Jin M, Deng Y, Wang JS, Huang H, Deng X, Huang CM (2019) Improvement of a deep learning algorithm for total electron content maps: image completion. J Geophys Res Space Phys 124(1):790–800
3. Choi BK, Sohn DH, Lee SJ (2019) Correlation between ionospheric TEC and the DCB stability of GNSS receivers from 2014 to 2016. Remote Sens 11(22):2657
4. Muafiry IN, Heki K, Maeda J (2018) 3D tomography of midlatitude sporadic-E in Japan from GNSS-TEC data. Earth Planets Space 70(1):45
5. Shi C, Zhang T, Wang C, Wang Z, Fan L (2019) Comparison of IRI-2016 model with IGS VTEC maps during low and high solar activity period. Results Phys 12:555–561
6. Li B, Wang M, Wang Y, Guo H (2019) Model assessment of GNSS-based regional TEC modeling: polynomial, trigonometric series, spherical harmonic and multi-surface function. Acta Geod Geophys 54(3):333–357
7. Zha J, Zhang B, Yuan Y, Zhang X, Li M (2019) Use of modified carrier-to-code leveling to analyze temperature dependence of multi-GNSS receiver DCB and to retrieve ionospheric TEC. GPS Solut 23(4):103
8. Sridhar KM, Sridhar M, Raghunath S, Ratnam DV (2020) Ionospheric anomaly detection and Indian ionospheric climatology from GAGAN receivers. Acta Geod Geoph 55:1–13
9. Tariku YA (2015) Comparison of GPS-TEC with IRI-2012 TEC over African equatorial and low latitude regions during the period of 2012–2013. Adv Space Res 56(8):1677–1685
10. Srivani I, Prasad GSV, Ratnam DV (2019) A deep learning-based approach to forecast ionospheric delays for GPS signals. IEEE Geosci Remote Sens Lett 16(8):1180–1184
11. Krypiak-Gregorczyk A (2019) Ionosphere response to three extreme events occurring near spring equinox in 2012, 2013 and 2015, observed by regional GNSS-TEC model. J Geod 93(7):931–951
12. Memarian Sorkhabi O, Asgari J, Amiri Simkooei A (2021) Analysis of Greenland mass changes based on GRACE four-dimensional wavelet decomposition. Remote Sens Lett 12(5):499–509
13. Gowtam VS, Tulasi Ram S, Reinisch B, Prajapati A (2019) A new artificial neural network-based global three-dimensional ionospheric model (ANNIM-3D) using long-term ionospheric observations: preliminary results. J Geophys Res Space Phys 124(6):4639–4657
14. Ferreira AA, Borges RA, Paparini C, Ciraolo L, Radicella SM (2017) Short-term estimation of GNSS TEC using a neural network model in Brazil. Adv Space Res 60(8):1765–1776
15. Sorkhabi OM, Asgari J, Amiri-Simkooei A (2021) Monitoring of Caspian Sea-level changes using deep learning-based 3D reconstruction of GRACE signal. Measurement 174:109004
16. Memarian Sorkhabi O (2015) Geoid determination based on log sigmoid function of artificial neural networks:(a case study: Iran). J Artif Intell Electr Eng 3(12):18–24
17. Liu C, Yu T (2019) The regulatory warning model of regional product quality based on the back-propagation artificial neural network. Neural Comput Appl 1–10
18. Valach F, Hejda P, Bochníček J (2007) Geoeffectiveness of XRA events associated with RSP II and/or RSP IV estimated using the artificial neural network. Stud Geophys Geod 51(4):551–562
19. Memarian Sorkhabi O, Asgari J, Amiri-Simkooei A (2021) Wavelet decomposition and deep learning of altimetry waveform retracking for Lake Urmia water level survey. Mar Georesour Geotechnol 1–11
20. Bishop CM (1995) Neural networks for pattern recognition. Oxford university press
21. Sorkhabi OM, Asgari J, Amiri-Simkooei A (2021) Monitoring of Caspian Sea-level changes using deep learning-based 3D reconstruction of GRACE signal. Measurement 174:109004
22. Gupta V, Singh JP (2019) Study and analysis of back-propagation approach in artificial neural network using HOG descriptor for real-time object classification. In: Soft computing: theories and applications. Springer, Singapore, pp 45–52
23. Djamour Y, Sorkhabi OM (2014) Estimation of geodetic velocity field based on sigmoid functions of artificial neural networks. Geospatial Eng J 5(4):49–55
24. Wen Z, Li S, Li L, Wu B, Fu J (2021) Ionospheric TEC prediction using long short-term memory deep learning network. Astrophys Space Sci 366(1):1–11
25. Yang H, Yuan C, Li B, Du Y, Xing J, Hu W, Maybank SJ (2019) Asymmetric 3d convolutional neural networks for action recognition. Pattern Recogn 85:1–12
26. Maru A, Dutta A, Kumar KV, Mohapatra DP (2020) Effective software fault localization using a back propagation neural network. In: Computational intelligence in data mining. Springer, Singapore, pp 513–526
27. Wan Q, Ma G, Li J, Wang X, Lu W, Maruyama T, Fan J, Zhang J (2020) Performance evaluation of IRI-2016 with GPS-derived TEC

at the meridian 110° E in China in 2014. J Atmos Solar-Terr Phys 105206

28. Mendoza LPO, Meza AM, Aragón Paz JM (2019) A multi-GNSS, multifrequency, and near-real-time ionospheric tec monitoring system for South America. Space Weather 17(5):654–661