Machine Learning-Based Short-Term GPS TEC Forecasting During High Solar Activity and Magnetic Storm Periods

Yi Han, Lei Wang, *Member, IEEE*, Wenju Fu, Haitao Zhou, Tao Li, and Ruizhi Chen

***Abstract*—Precise ionospheric total electron content (TEC) is critical for many aerospace applications, and forecasting ionospheric TEC is of great significance to it. Besides, short-term prediction of TEC values fills the gap between the TEC product latency and the precision. The machine learning-based approaches are promising in solving the nonlinear prediction issues, particularly suitable for short-term global positioning system TEC forecasting due to its complex temporal and spatial variation. In this article, four different machine learning models, i.e., artificial neural network, long short-term memory networks, adaptive neuro-fuzzy inference system based on subtractive clustering, and gradient boosting decision tree (GBDT) are applied for forecasting ionospheric TEC in three IGS GNSS monitoring stations at the low-latitude region (16°S to 10°S). The performance of these approaches in extreme conditions is investigated, including the high solar activity and magnetic storm, which are the most challenging scenario for TEC prediction. The results show that the machine learning algorithms outperform the global ionospheric map prediction model. The prediction accuracy during the high solar activity period was improved from 37.93% to 49.28%. During the magnetic storm period, the prediction accuracy was improved from 28.16% to 67.39%. Among the machine learning algorithms, the GBDT model outperforms the rest three algorithms in ionosphere prediction scenarios, which improves the prediction accuracy by 5.6% and 12.7% than the rest three approaches on average during high solar activity (2012–2015) and magnetic storm periods respectively.**

***Index Terms*—Gradient boosting decision tree (GBDT), ionosphere prediction, machine learning, neural network (NN), total electron content (TEC).**

I. Introduction

CONTINUOUS monitoring of the ionosphere layer has been carried out using many geodesy techniques due to significant effects on both communications and global navigation satellite systems (GNSSs) [1]–[4]. The ionosphere changes fast and complex over different times and places, which limited the precise ionosphere acquisition capacity. The ionosphere working group of the International GNSS Service (IGS) has been continuously providing reliable global ionospheric maps (GIMs) since 1998 [5]. The nominal precision for the final product is about 2-8 TECU, while it comes with about 3-day latency [6]. The Center for Orbit Determination in Europe (CODE) also published the prediction product C1PG and C2PG for real-time applications [7]. However, its prediction accuracy largely depends on solar activity. During the high solar activity period, the ionosphere prediction is still challenging and the C1PG accuracy is about 4-16 TECU [8]. In addition, International Reference Ionosphere 2016 (IRI-2016) model is a standard empirical model, whose goal is to develop and improve international standards for the parameters of the Earth's ionosphere [9]. In low-latitudes, the accuracy of IRI-2016 model is about 7-15 TECU during high solar activity period, when compared to IGS GIM [10]. Thus, precise TEC prediction in the extreme conditions is a big new challenge to deal with.

The ionosphere prediction is an important way to obtain the real-time ionosphere TEC and there have been a few researches on this topic using the classical time-series analysis approaches, such as the auto-regressive moving average (ARMA) model and auto-regressive integrated moving average (ARIMA) model are widely applied for short-term TEC forecasting in different latitudes [11]–[13]. The machine learning algorithms can automatically learn the implicit nonlinear relationship between the TEC value and the external indicators, which helps to improve the prediction performance in extreme environments. A few machine learning approaches have been used in the ionosphere prediction, such as the standard neural network (NN) approach [14]–[16], long short-term memory (LSTM) [17]–[19], adaptive neuro-fuzzy inference system (ANFIS) [20], etc. Their performance has been evaluated and has been compared with the GIM and IRI-2016 models and the results show that these machine learning algorithms outperform the existing models. However, the performance between different machine-learning algorithms has not been evaluated so far, and the optimal machine learning algorithms have not been identified.

In this study, a new machine learning algorithm named gradient boosting decision tree (GBDT) was introduced for the ionosphere prediction problem. Four machine learning algorithms are selected, such as the NN approach, the LSTM approach, the

Manuscript received March 26, 2021; revised August 23, 2021; accepted November 24, 2021. Date of publication December 2, 2021; date of current version December 22, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant NSFC 42074036 and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding author: Lei Wang.)*

The authors are with the State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China, and also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: yi_han@whu.edu.cn; lei.wang@whu.edu.cn; wenjufu@whu.edu.cn; haitao.zhou@whu.edu.cn; tao.li@whu.edu.cn; ruizhi.chen@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3132049

ANFIS approach, and the GBDT approach, to compare their performances in extreme conditions, which aim to find the best machine learning algorithm during the high solar activity and magnetic storm period. Three monitoring stations from the IGS network at low-latitude regions were selected, and four-year data (2012–2015) during the high solar activity and magnetic storm period (46 d) are used in this study. Then, many accuracy indicators to compare their overall performance are chosen in detail.

The rest of this article is organized as follows. Section II describes the vertical TEC (VTEC) values calculation, and the models used in the TEC prediction are presented in Section III. In Section IV, experiment setups are presented. The comparisons of the nonlinear models and results with related researches are discussed in Section V. Finally, Section VI concludes this article.

II. VTEC Extraction From GNSS Observations

To accurately forecast the TEC time series, the TEC values need to be extracted from dual-frequency GNSS receivers. The code and phase of the carrier wave observations are provided by the dual-frequency global positioning system (GPS) receivers in the L band at frequencies of L1 (1575.42 MHz) and L2 (1227.60 MHz), and these observations are most used to estimate the slant TEC (STEC). The ionosphere time delays can be extracted from pseudoranges and carrier phase observations, respectively. The GNSS observations can be expressed as follows:

$$\begin{aligned} P_i &= D + T + I_i + \mathrm{b}_i^s + b_{r,i} + c(\Delta t^s - \Delta t_\mathrm{r}) + \varepsilon_P \\ \varphi_i &= \frac{f_i}{c}(D^s + T^s - I_i^s + B_i^s + B_{r,i} + c(\Delta t^s - \Delta t_\mathrm{r})) \\ &\quad + N_i + \varepsilon_\phi \end{aligned} \tag{1}$$

where denotes pseudorange between receiver r and satellite s at the ith frequency; $D$is the geometric distance between the satellite and receiver; T, Irepresents the tropospheric path delay and the ionospheric delay; Δt^s and Δt_r are the P_r^ssatellite-specified and receiver-specified clock biases; b_i^s and $b_{r,i}$ are the satellite-specified and receiver-specified code biases for ith frequency; B_i^s and $B_{r,i}$ are the satellite-specified and receiver-specified uncalibrated initial fractional carrier phase biases. $\varphi$is the carrier phase observations expressed in the cycle; $N$is the carrier phase ambiguity bias. f_i and c are the frequency of the observation and the light speed, respectively.

The ionospheric delay in GNSS observations is frequency-dependent due to ionospheric dispersive nature; hence, the total GPS ionosphere delay can be estimated with dual-frequency GNSS observations. The STEC can be derived according to the total ionosphere delay. GPS STEC can be estimated with either the phase smoothing pseudorange method or the precise point positioning (PPP) method [21], [22]. The phase smoothing pseudorange method is the most widely used method, which is also used to generate the global ionospheric map product; hence, we adopt this method to estimate the STEC. This STEC can be derived from the following equation:

$$\begin{aligned} P_4 &= P_1 - P_2 \\ L_4 &= \lambda_1\phi_1 - \lambda_2\phi_2 \\ \mathrm{STEC} &= \frac{f_1^2 f_2^2}{40.3\,(f_2^2 - f_1^2)}\,(P_4 - c\cdot \mathrm{DCB}_r - c\cdot \mathrm{DCB}^s) \end{aligned} \tag{2}$$

where P_4 and L_4 are the geometry-free code and carrier phase combination, respectively, which canceled all the frequency-independent terms in (1). The remaining frequency-dependent terms can be further decomposed into three parts: the STEC, the satellite-specified, and receiver-specified differential code biases (DCB). DCB is the interfrequency code biases. For dual-frequency case, DCB can be defined as $\mathrm{DCB}_r = b_{r,1} - b_{r,2}$, $\mathrm{DCB}^s = b_1^s - b_2^s$. In order to mitigate the pseudorange noise impact in STEC estimation, a carrier phase smoothed pseudorange observations is adopted, given as follows:

$$\tilde{P}_4^k = L_4^k - \frac{1}{n}\sum_{k=1}^{n}(L_4^k + P_4^k) \tag{3}$$

where $\tilde{P}_4$ is the carrier phase smoothed pseudorange observations, n is the number of epochs. According to the ionosphere thin shell model, the STEC can be converted to VTEC at the ionosphere pierce point (IPP). Finally, the TEC in the zenith direction can be calculated as follows:

$$\mathrm{VTEC} = \cos(Z') \times \mathrm{STEC}\ \text{with}\ \sin(Z') = \frac{R}{R+H}\sin Z \tag{4}$$

where R is the mean Earth radius, Z' and Z are the zenith angles of the satellite at the receiver and at the ionospheric pierce point, respectively. The estimated VTEC of 1 h resolution is considered as the dataset for machine learning models, which are separated into training and testing datasets. The three low-latitude IGS stations are selected for this study. Besides, a training dataset is used to train models and the k-fold validation method is also applied [23]. This step can estimate the generalization ability of the machine learning models, and can be used to identify whether the network is overfitting and underfitting. The testing dataset is used to evaluate the predicting capacity of these models.

III. Machine Learning Models for TEC Prediction

In this section, the four machine learning methods and their usage are presented.

A. Artificial NNs

Artificial NNs are well known as a powerful and elegant technique to approximate a complex nonlinear function as a composition of elementary nonlinear functions [24]. This model has been successfully applied in TEC modeling and forecasting. The NN establishes the nonlinear relationship between the inputs and the outputs by adjusting the weight between the nodes of the network. A trained network contains the implicit nonlinear relationship between the input and the output, and can be used for the ionosphere prediction with only input features. Taking X as the input vector in the NN, the output of the network can be

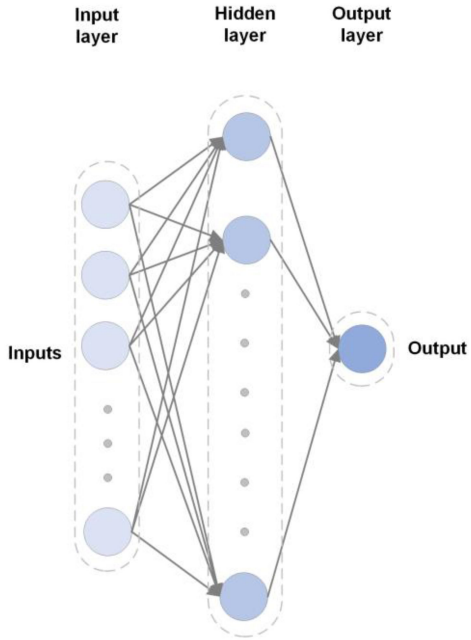


Fig. 1. NN structure with one hidden layer.

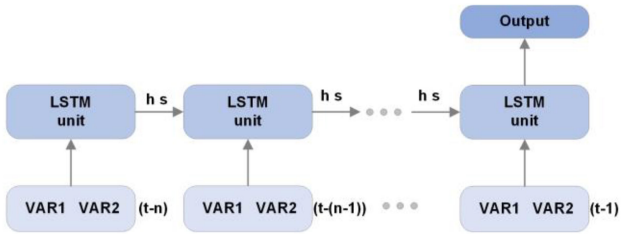


Fig. 2. Architecture of the LSTM model.

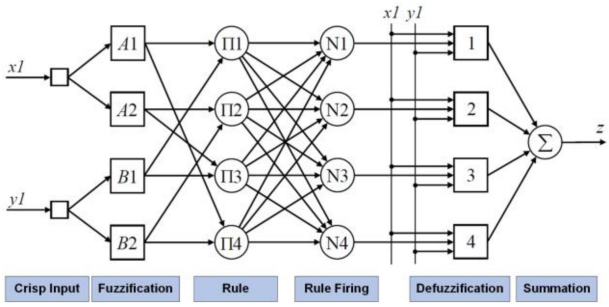


Fig. 3. ANFIS structure with two inputs.

expressed as follows:

$$y = k\left(\sum_{j=1}^{n}(w_j x_j + b)\right) \tag{5}$$

where $k()$ is the activation function; $w$denotes the weight of each neuron, and $b$is the bias; $x$is the input feature. Particularly, activation function determines the relationship between inputs and outputs of a NN, which introduces a certain degree of nonlinearity. In this study, rectified linear unit (ReLU) is used as activation function [25]. An illustration of the NN structure is presented in Fig. 1.

In TEC forecasting, X is the input features related to VTEC value, and y is the predicted TEC value. The backpropagation algorithm is chosen in this study to construct the NN model [26].

B. LSTM Networks

The single station VTEC prediction can be taken as the time series prediction problem. The LSTM network is a specially designed network for sequential data processing [27]. LSTM has the capacity of learning the long-term dependency and the short-term dependency at the same time. With new input data accumulated, the network is evolutionary, while the long-term dependency was preserved with the hidden state vector, so that the predicted data can benefit from both historical and the latest input data. The flowchart of LSTM is shown in Fig. 2.

The output of LSTM is presented as follows [28]:

$$h_t = o_t \tanh(c_t) \tag{6}$$

where h_t represents the hidden state vector as well as the output vector of the LSTM unit; o_t is the activation vector of output gate, and c_t is the cell state vector. VAR1 and VAR2 are the input features related to VTEC value in the LSTM model, which uses the sequential data to predict 1 h ahead TEC values. See [17] and [27] for details about algorithms and their usage.

C. Adaptive Neuro-Fuzzy Inference System

ANFIS is a combination of NNs and the fuzzy inference system (FIS), which gains their advantages of using numerical and linguistic power [29]. Since it introduces fuzziness into the NN, the combination of this model can increase the accuracy of results. Existing research reveals that the ANFIS shows better performance than the classical NN in TEC modeling at the low solar activity [30]. Thus, this model is most likely to be used in TEC forecasting under severe weather conditions in terms of its fuzziness characteristic. The simple structure of an ANFIS network is shown in Fig. 3.

Like the NN, ANFIS consists of at least five layers. In the first layer, the two crisp inputs are put in this system, and they are fuzzified into values by a fuzzy set described as neurons, which has its membership function with adjustable premise parameters. In the rule layer, each node in this layer receives signals from the fuzzification layer, while the degree of activation of the premises is calculated. Each node of this layer represents the firing strength of a fuzzy rule. The normalization is done in this rule firing layer. In the next layer, the adjusted consequent parameters are contained in each neuron. Lastly, the summation of received signals of previous nodes is calculated in this layer as the final output value of the ANFIS system.

According to the number of input variables, the ANFIS can be established using different methods, which include grid partitioning and clustering techniques [31]. Regarding the grid partitioning method, as the number of input variables increases, the number of fuzzy rules increases exponentially. Therefore,

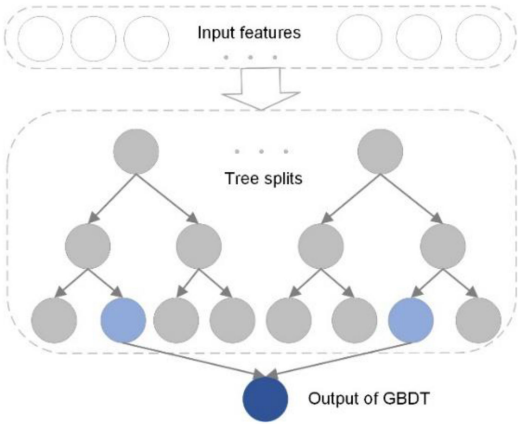


Fig. 4. Illustration of the GBDT model structure.

large fuzzy rules pose a threat to the application of ANFIS, and this method is suitable for datasets with a small number of input variables (e.g., less than 6) [32]. However, the inputs in this work are more than six, and so this grid method is not considered. To enable this system to be able to solve the multi-input problem, the clustering technique is applied as an effective tool to classify the inputs to reduce the fuzzy rules, while the number of clusters is equal to the fuzzy rules. So, in this study, the ANFIS subtractive clustering (ANFIS-SC) network is used to establish forecasting model.

D. Gradient Boosting Decision Tree (GBDT)

The GBDT approach combines regression trees with a gradient boosting technique, which has been widely used to solve regression and classification problems such as soil moisture estimation [33], GNSS positioning accuracy improvement [34], and GPS signal reception classification [35]. It achieves the best performance when comparing with the other classic machine learning techniques [35], [36]. The GBDT approach is high efficiency and has an excellent capacity to automatically detect nonlinear interactions, which is quite promising for accurately forecast TEC time series under severe weather conditions.

The detail of this algorithm is described as follows. In this algorithm, each sample is represented as x_i $(i = 1, 2, 3, \ldots, N)$, and N is the number of samples. The training set can be expressed as $T = \{(x_1, y_1)(x_2, y_2), \ldots, (x_N, y_N)\}$, while y is the TEC predicted value of each sample. The GBDT minimizes the value of some specified loss function to obtain the optimal model, and loss function is used to obtain residuals at each iteration [37]. The most used function is the square function and absolute error. In this study, the square loss function $L(y_i, f(x_i)) = \frac{1}{2}(y_i - f(x_i))^2$ is used, and the GBDT model structure is shown in Fig. 4.

The main steps of the method are as follows [35], [37]:

Algorithm

1) For the training set, initialize a weak learner $f_0(x)$:
   $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$
   $f_0(x)$ is a regression tree, and according to the square loss function, $f_0(x)$ becomes:
   $f_0(x) = \overline{y}$
2) For $m = 1$ to M :
a) Compute the negative gradient
   $r_{mi} = -[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}]_{f=f_{m-1}}$
b) Replace y_i of the training dataset with r_{mi} to obtain a new dataset $\{(x_1, r_{m1})(x_2, r_{m2}), \ldots, (x_N, r_{mN})\}$. $f_m(x)$is the new regression tree, and $R_{mj}$is its areas of leaf nodes, where $j = 1, 2, 3 \ldots, J$. Jdenotes the number of leaves in a given regression tree, and calculate the optimal fitting values:
   $\gamma_{mj} = \arg\min_\gamma \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + \gamma)$
c) Update the strong learner:
   $f_m(x) = f_{m-1}(x) + \rho \sum_{j=1}^{J_m} \gamma_{mj} I(x \in R_{mj})$
where $\rho$is the learning rate to prevent overfitting issues.
3) When the iteration is terminated, the final output becomes:
   $f_M(x) = f_0(x) + \rho \sum_{m=1}^{M} \sum_{j=1}^{J_m} \gamma_{mj} I(x \in T_{mj})$

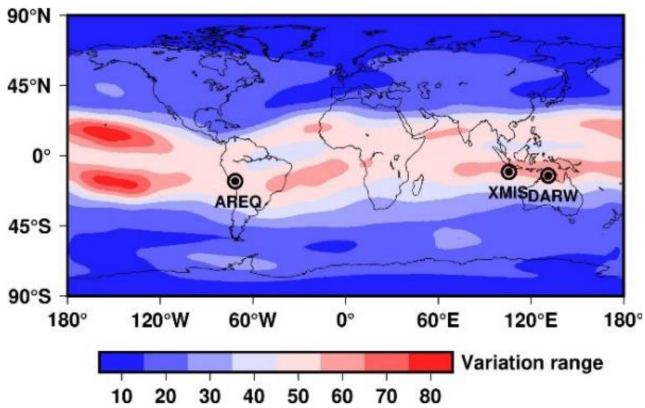


Fig. 5. Distribution of the IGS stations and the ionosphere variation range.

IV. Experiments Setup

To evaluate the forecasting performance of these models, several experiments are to be performed. In this section, the description of the experiment setup, which includes the description of the datasets, evaluation metrics and parameters setting are introduced.

A. Datasets

In order to capture the ionosphere variation characteristics, the dataset used in this study is carefully selected. At first, three continuous monitoring stations from IGS network are selected. As known, the ionosphere changes the most dramatically in the low latitude belt region around the magnetic equator. We checked the mean daily variation of the global TEC using the GIM products from CODE and then three IGS stations at the low magnetic latitude was selected as the target. Th mean daily TEC variation and selected IGS stations are presented in Fig. 5.

TABLE I
TEC AND FEATURE CHARACTERISTICS IN 2012–2015

Dataset	Station		Maximum	Minimum	Median	Mean	Std
TEC	AREQ	Training set	108.5974	1.4303	33.3057	35.8694	19.7417
(TECU)	(16.466°S, 71.493°W)	Testing set	101.9865	2.4708	31.3346	34.1138	19.3158
	DARW	Training set	124.3451	4.8	28.2067	31.5464	17.6207
	(12.844°S, 133.133°E)	Testing set	111.4639	3.9733	23.4522	27.5892	16.6845
	XMIS	Training set	124.1335	0.0362	29.0292	33.3563	23.025
	(10.450°S, 105.688°E)	Testing set	120.0678	0.3405	24.7426	29.7139	21.9752
F10.7		Training set	253.3	86.8	127.1	129.4493	24.9712
		Testing set	255	79.5	115.5	117.6025	19.8364
Kp		Training set	8	0	1.3	1.5779	1.231
		Testing set	8.3	0	2	2.1075	1.375
Dst		Training set	77	-145	-8	-10.9685	17.961
		Testing set	56	-223	-10	-14.31	21.6968

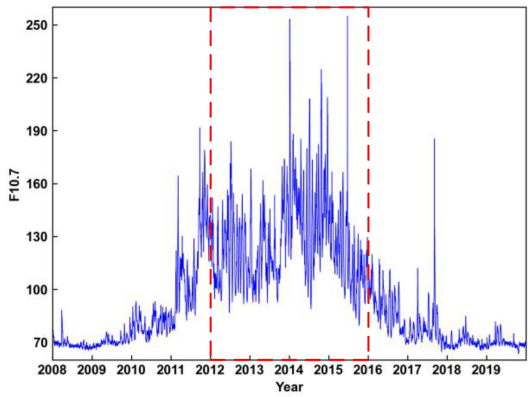


Fig. 6. F10.7 value time series for the recent 12 years.

In the temporal dimension, the solar activity has a period of about 11 years and its activity can be reflected with the F10.7 index. We checked the F10.7 index time series for the recent 12 years and selected the period 2012 to 2015 as the target since this period has the highest solar activity in the recent 12 years, which is shown in Fig. 6. In practice, the data from 2012 to 2014 are used for model training and the data from 2015 are used to testing the ionosphere prediction performance.

For the machine learning-based algorithm, the selection of the features is quite critical. The variation of TEC is associated with the season, day, hour, geomagnetic activity, and solar activity. The solar zenith angle (SZA) contains characteristics of season, day, and hour. In addition, the geomagnetic activity and solar activity can be characterized by the Dst, Kp, and F10.7 index [38], [39]. In this study, five features are considered, namely, SZA, F10.7, Dst, Kp, and TEC value. The detailed statistical characteristics of these datasets are listed in Table I. It should be mentioned that the variation and range of the SZA dataset is periodically variation in the interval $[-90, 90]$, and so its statistical data is no need to be presented. The TEC variation in the selected period is up to 120 TECU for the three IGS stations, which can be considered as a typical high solar activity period. The time series for the selected features over the four years are presented in Fig. 7. The figure gives an overall description of the data. The VTEC value time series on the three stations is presented in the figure.

It indicates that the VTEC has certain periodical patterns, and two peaks are observed each year at this station. F10.7 index is a

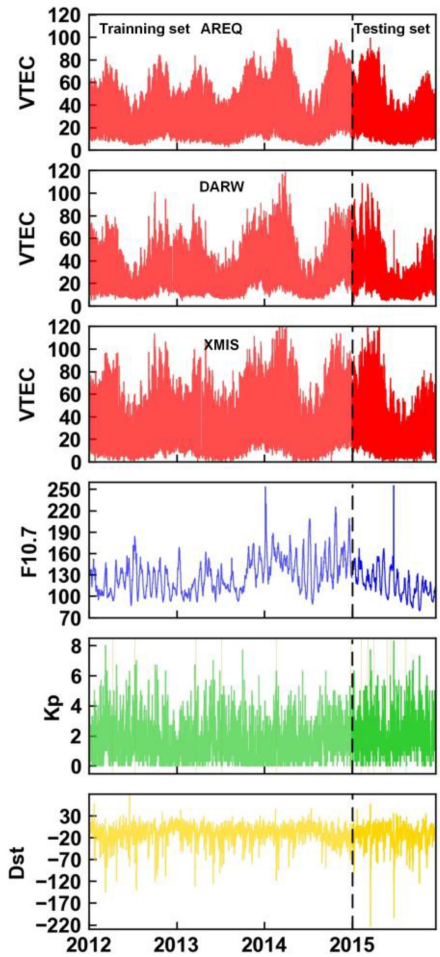


Fig. 7. TEC and feature time series in 2012–2015.

measure of the noise level generated by the sun at a wavelength of 10.7 cm at the earth's orbit. A higher F10.7 value means more active ionosphere activity [40], [41]. The Kp and Dst indexes are used as indicators of the geomagnetic activity. Kp index presents the index of 3-h range in magnetic activity relative to a quiet day. Thus, a higher Kp index means the geomagnetic is more active. Generally, $Kp \geq 4$ means high geomagnetic activity. The Dst index smaller -50 nT can be considered as the moderate storms.

TABLE II
DEFINITION OF THE PERFORMANCE METRICS

Criteria	Definition	Equation
RMSE	Root mean square value of the errors	$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$
MAE	Mean absolute error	$MAE = \frac{1}{N} \sum_{i=1}^{N} \lvert y_i - \hat{y}_i \rvert$
MAPE	Mean absolute percentage error	$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left\lvert \frac{y_i - \hat{y}_i}{y_i} \right\rvert \times 100\%$
TIC	Theil inequality coefficient	$\mathrm{TIC} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \Big/ \left(\sqrt{\frac{1}{N} \times \sum_{i=1}^{N} y_i^2} + \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \hat{y}_i^2} \right)$
R^2	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$
Var	Variance of the forecasting error	$Var = E(\mathbf{e} - E(\mathbf{e}))^2$

B. Performance Metrics

To evaluate the performance of these machine-learning-based models, different performance metrics are considered in this study. The prediction quality of a model is generally examined with two aspects of the prediction residuals. 1) Bias, which is the deviation from the true value distribution. 2) Variance, which reflects the stability and the robustness of the model prediction capacity.

It should be noticed that single criteria cannot fully measure the performance of these forecasting models. Hence, a comprehensive evaluation system should compose of more performance indices, which includes the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the Theil inequality coefficient (TIC), the correlation coefficient (CC), and forecasting error variance (VAR) is used in this article.

Based on the forecasting and actual value, the RMSE, MAE, MAPE, and TIC are calculated to assess the forecasting accuracy. The RMSE can efficiently evaluate the forecasting accuracy. In addition, since the forecasting error can be positive and negative, the MAE is the average of the absolute error, which can better reflect the actual situation of the predicting value error, and MAPE is applied to measure the relative error between the average test value and the actual value. The TIC is calculated using forecasting error, while the lower the TIC, the better forecasting performance. Besides, R^2 indicates the degree of linear correlation between forecast value and actual value. The VAR is applied to evaluate the nonlinear model stability. Definitions of these metrics are listed in Table II.

C. Configuration of the Machine Learning-Based Models

The performance of the machine learning-based models depends on the configurations of the models. Typical machine learning procedures involve a training procedure and a testing procedure. The training procedure extracts the implicit relationship using the input features and the "true" TEC values. Then the trained network is used to predict the TEC value with given input features. The quality of the training process is measured by the RMSE of the fitting residuals. However, the smallest RMSE of the fitting residuals does not necessarily imply better prediction accuracy due to the overfitting issue. In this study, 75% of the training dataset is used for training the models and the rest 25% are used for testing the model. In the training process, a five-fold validation method is used to avoid the overfitting problem [42]. The training set is randomly divided into five groups equally. Four groups are involved in the training and the rest group is used for the validation procedure in each round and five rounds of the training process are carried out until all groups are used as the validation group to find the optimal intrinsic parameters. Then, the rest 25% testing dataset is used to further evaluate the predicting performance of these models. Since the testing dataset was not involved in the training procedure, it is reasonable to evaluate the performance of the machine learning-based approach with the testing dataset.

The parameter setting for these models is described as follows. In NN and LSTM models, the number of neurons is between 1 and 100, and the learning rate is within 0.005 and 0.05. Their hidden layers are set to 1, while the iterations are set to 1000. For ANFIS-SC models, the range of influence is between 0.4 and 0.6, and iteration is set to 100. Finally, for GBDT models, the learning rate is between 0.01 and 0.1, and the number of estimators is between 500 and 1500. The number of depths is between 3 and 9. All parameter definitions for each machine learning model are summarized in Table III.

V. PERFORMANCE EVALUATION

A. Optimal Window Length Determination

In the machine learning-based approaches, the VTEC for the next hour is modeled as a function of the features in the past few hours, which is also known as the window length of the ionosphere prediction. In this study, number of hours ahead prediction of ionospheric TEC is one hour, and window length represents the data of the previous few hours.

The window length has an impact on the prediction precision, so how to optimally determine the window length should be investigated first. We use the MAPE indicator as the prediction performance measure and compared different window lengths using the four machine learning algorithms with the data from

TABLE III
INTRINSIC PARAMETERS USED IN EACH MACHINE LEARNING-BASED MODELS

Station	Model	Main parameter definitions
AREQ	NN	learning rate=0.005, number of neurons=80
	LSTM	learning rate=0.01, number of neurons=98
	ANFIS&SC	Range of influence=0.4
	GBDT	learning rate=0.07, n_estimators=109, max_depth=14
DARW	NN	learning rate=0.005, number of neurons=97
	LSTM	learning rate=0.005, number of neurons=97
	ANFIS&SC	Range of influence=0.4
	GBDT	learning rate=0.03, n_estimators=1310, max_depth=4
XMIS	NN	learning rate=0.005, number of neurons=96
	LSTM	learning rate=0.01, number of neurons=83
	ANFIS&SC	Range of influence=0.4
	GBDT	learning rate=0.01, n_estimators=1500, max_depth=7

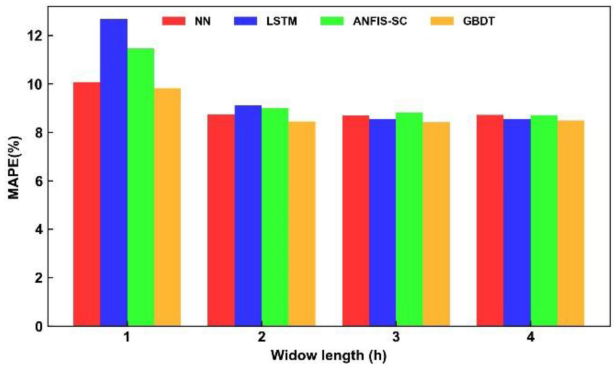


Fig. 8. MAPE values for different window length in AREQ station.

the AREQ station, and the results are presented in Fig. 8. The results indicate that different machine learning algorithms have different window length dependencies and the LSTM presents the strongest window length dependency. It also shows that as the window length increases, the MAPE becomes smaller in all models. Generally, longer window length implies better prediction accuracy, while the improvement of window length longer than 3 h becomes negligible. In order to balance the prediction precision and the computation complexity, 3 h window length is used in this study for the ionosphere prediction.

B. Performance Comparison Evaluation in High Solar Activity Period

In order to evaluate the prediction performance of machine learning-based approaches in the high solar activity period, we tested their performance with the TEC time series derived from the three stations (AREQ, DARW, and XMIS) with 1 h temporal resolution from 2012–2015.

In this study, the data in 2012–2014 is used for model training and the data in 2015 is used as the testing dataset. As stated, this period is the highest solar activity period in recent 12 years. The ionosphere prediction results of these models are listed in Table IV. For the AREQ station, the MAPE values vary from 8.4% to 8.7% for the four machine learning models, and vary from 13.6% to 28.7%, which is shown in Table IV.

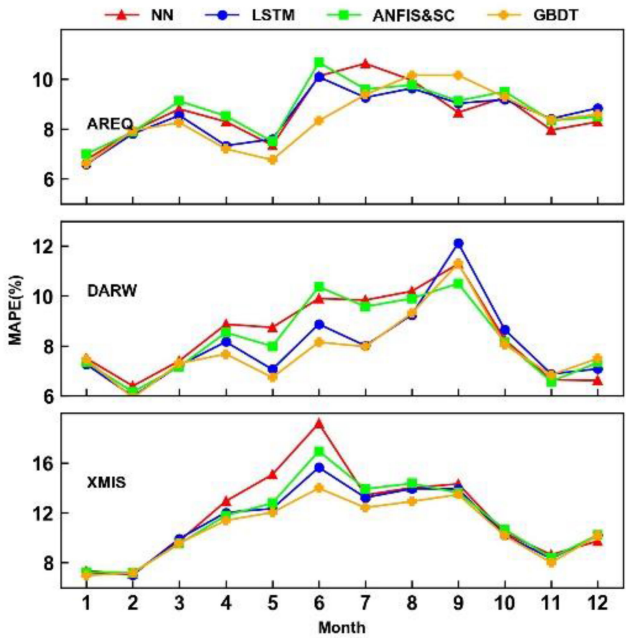


Fig. 9. Monthly MAPE values of the four machine learning-based models.

The prediction accuracy of the four machine learning-based approaches is about 8%, while the rest two global models are higher than 10%. The table also indicates that the GIM model achieves better accuracy than the IRI-2016 model in the highest solar activity period. The four machine learning-based models outperform the GIM and IRI-2016 models in terms of other performance indicators as well.

In addition, the GBDT model also outperforms the other models in all stations in terms of the rest performance indicators. Particularly, the smallest VAR indicator indicates that the GBDT also achieves the most stable prediction performance over the high solar activity period. In terms of the RMSE, the prediction accuracy of the GBDT model is better than the rest three models by 5.6% on average. Thus, it can be concluded that the GBDT model outperforms the rest three machine learning-based models in the high solar activity period, although some researchers have achieved fairly good ionosphere prediction accuracy in relatively low solar activity period [17]. The LSTM may not be the best choice for ionosphere prediction in the high solar activity period.

Even in the high solar activity period, the ionosphere activity also varies with the season. In order to investigate the impact of season on the machine learning-based approaches, we analyzed the monthly mean MAPE of the four approaches and the results are presented in Fig. 9. The figure shows that the MAPE of the four approaches has different characteristics. Generally, the four approaches perform similarly from October to next February, while the discrepancy increases from March to September. The performance of the different approaches is also station specified, but the GBDT achieves the smallest MAPE in most cases. Although the station DRAW and XMIS are geographically close to each other, their prediction error did not present similar behavior. The largest discrepancy between the four models presents in June for all three stations and the

TABLE IV
OVERALL PERFORMANCE OF THE FOUR MACHINE LEARNING-BASED ALGORITHMS DURING THE TESTING PERIOD

Station		RMSE	MAE	MAPE (%)	R^2	TIC	VAR
AREQ	GIM	5.2694	3.7972	13.6394	0.9259	0.066	26.5263
	IRI-2016	13.8511	10.2522	28.7071	0.4879	0.2062	111.1209
	NN	3.3401	2.3446	8.6769	0.9702	0.0427	11.1557
	LSTM	3.3024	2.3244	8.5387	0.9709	0.0423	10.9032
	ANFIS-SC	3.4658	2.386	8.8061	0.9679	0.0443	12.0079
	GBDT	**3.2707**	**2.2887**	**8.4144**	**0.9714**	**0.0418**	**10.6851**
DRAW	GIM	5.7268	3.8478	15.849	0.8823	0.0891	32.7513
	IRI-2016	10.8644	8.0071	30.7649	0.5762	0.1854	95.5641
	NN	3.0022	2.0697	8.4973	0.9676	0.0467	9.0107
	LSTM	3.0236	2.015	8.1757	0.9672	0.0472	9.1358
	ANFIS-SC	3.4079	2.0222	8.3817	0.9583	0.0529	11.5852
	GBDT	**2.9757**	**1.9761**	**7.9992**	**0.9682**	**0.0461**	**8.7672**
XMIS	GIM	7.4888	5.1338	23.9058	0.8844	0.1019	56.0809
	IRI-2016	14.2453	9.6448	36.5121	0.5818	0.2181	175.814
	NN	3.8383	2.5922	11.8452	0.9696	0.0518	14.6144
	LSTM	4.0301	2.634	11.209	0.9665	0.0546	16.1853
	ANFIS-SC	4.7503	2.9298	12.8272	0.9535	0.0643	22.5436
	GBDT	**3.7979**	**2.5157**	**10.7162**	**0.9703**	**0.0514**	**14.4005**

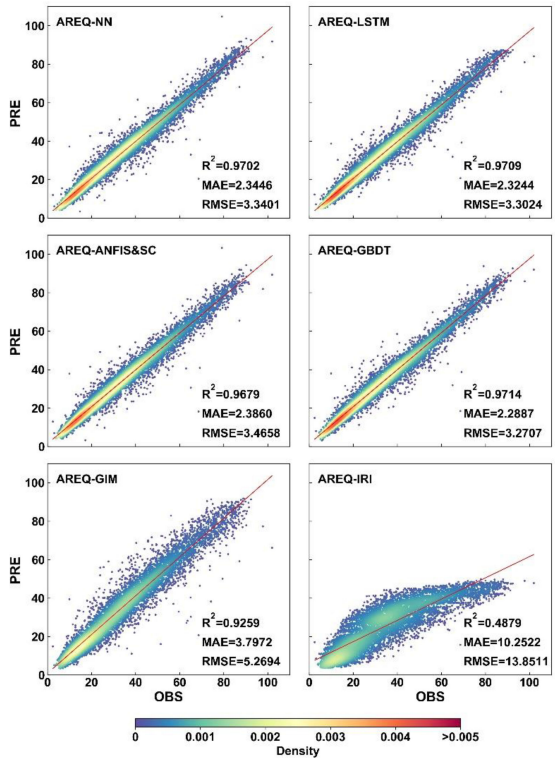


Fig. 10. Correlation between observed GPS TEC and predicted TEC value using the four machine learning-based models during the 2016 period.

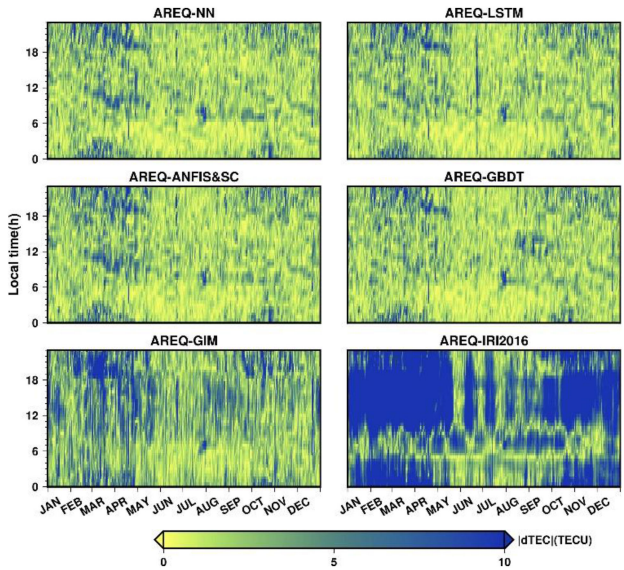


Fig. 11. Relationship between the absolute |dTEC| prediction error and the season/local time in the AREQ station.

GBDT approach did not perform well in September for both AREQ and DRAW. Generally, the ANFIS and NN perform not as good as the LSTM and GBDT. Although the LSTM performs well, GBDT still slightly outperforms, especially in the middle of the year.

The correlation between the predicted VTEC value and the observations for all six approaches is also examined and the results are presented in Fig. 10. The figure shows the four machine learning-based approaches can perfectly capture the variation of the ionosphere VTEC and the R^2 values of the four models are generally higher than 0.97. The GIM model also presents perfect correlation, but its precision is still not as good as the machine learning-based approach. The IRI-2016 model presents certain biases in the ionosphere VTEC computation and also presents the poorest correlation. The figure also indicates that the ionosphere VTEC predicted with the IRI-2016 model presents a strong bias in high VTEC scenarios, which indicates the model still can be improved to cope with the high solar activity scenarios.

The ionosphere activity is also related to the SZA, and hence we further investigated the relationship between local time and

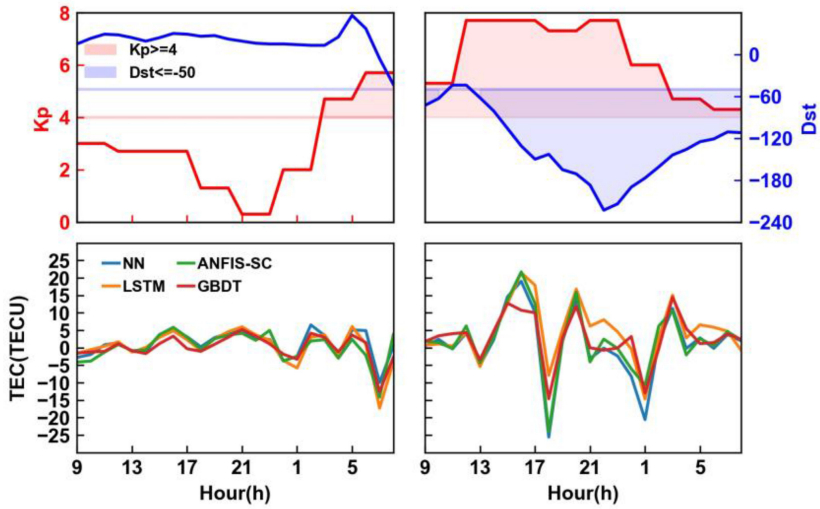


Fig. 12. Ionosphere prediction accuracy comparison of the four machine learning-based algorithms in magnetic peace (left) and storm (right) period during days 75–77.

the ionosphere prediction error. The |dTEC| distribution for the AREQ station in 2015 is presented in Fig. 11.

This figure shows the prediction error is related to the local time and the season for all six approaches. The ionosphere prediction error is smallest from June to September and the prediction error achieves the maximum during February to May. In the first half-year, there are two prediction error peaks presented in each day, while the prediction error is more homogeneous in the second half-year. Generally, all these approaches achieve the smallest prediction error in the night (Local time 0–6 h) due to relatively peaceful ionosphere activity. The prediction error for the four machine learning-based approaches are very similar, which indicates the ionosphere VTEC can be quite unpredictable at some times.

C. Performance Evaluation in the Geomagnetic Storm Period

A more extreme weather condition for the ionosphere prediction is the geomagnetic storm. The ionosphere activity is strongly correlated to the magnetic disturbance, which may be affected by many external factors. The magnetic storm may cause power outages, satellite communication failure, and satellite damages, and so it is taken as hazardous space weather [43], [44]. In this study, we also introduce two magnetic disturbance indices to help predict the ionosphere VTEC in the model training phase. The geomagnetic storm is empirically defined as $\mathrm{Dst} <= -50$ nT and $\mathrm{Kp} >= 4$ in this study [45]. Different from solar activities, a geomagnetic storm is a short-term event, which typically lasts a few hours to several days. To investigate the performance of the machine learning-based algorithms in the magnetic storm, we selected 46 d in 2015, which meet the magnetic storm criteria.

We compared the impact of the magnetic storm on the ionosphere prediction and the results are presented in Fig. 12. The figure compared the VTEC value and the prediction error in the magnetic peace and storm period. The magnetic storm happened on the day-of-year (DOY) 077, 2015 and the right panel shows the maximum VTEC reaches 100 TECU in the storm. The left panel shows the VTEC and prediction error at the same time as the previous day. Although the magnetic storm has not been started, the maximum VTEC also reaches about 70 TECU. The figure indicates the magnetic storm has a significant impact on ionosphere prediction. The prediction error increases from a few TECU to about 20 TECU during the magnetic storm. In the magnetic peace period, all the machine learning approaches perform quite similarly, but in the storm, the ionosphere becomes quite unpredictable. Within the four approaches, the LSTM and the ANFIS approaches are more apt to affect by the storm. The NN and GBDT perform slightly better than the other two approaches.

We further investigated the impact of the magnetic storm on the ionosphere TEC prediction residuals and the results on AREQ are presented in Fig. 13. The figure shows that the impact of the magnetic storm is obvious. The storm leads to larger prediction residuals for all six approaches and the RMSE of the prediction residuals is increased by 2–4 TECU. The RMSE, MAE, and R2 values in the storm time are all larger than those in nonstorm time. Particularly, the magnetic storm increases the probability of large prediction residuals, which makes the ionosphere prediction in the magnetic storm less reliable. Comparing the four machine learning-based approaches, it concludes that the GBDT outperforms the rest three approaches in terms of R2, MAE, and RMSE during the magnetic storm period. While the superiority becomes less significant in the nonmagnetic storm period, it also performs well.

The overall performance of the six approaches during the geomagnetic storm is compared and the results are listed in Table V. The table shows that the prediction accuracy of all six models is lower than that of the whole high solar activity period. The main reason for this difference may be that the TEC variation becomes more complex and unpredicted during the geomagnetic storm period. Specifically, the MAPE values of four machine learning models for forecasting GPS TEC time series are all smaller than GIM and IRI-2016 models. The GBDT and NN

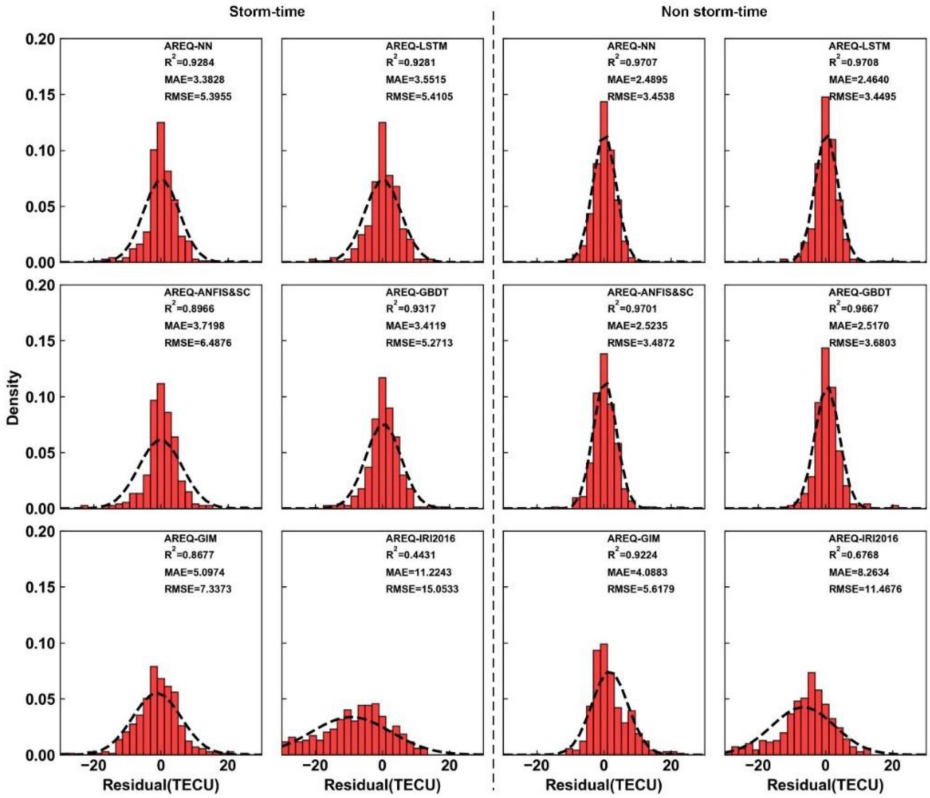


Fig. 13. TEC prediction residuals comparison during the magnetic storm and nonstorm period.

TABLE V
OVERALL PERFORMANCE OF THE IONOSPHERE PREDICTION MODELS DURING THE GEOMAGNETIC STORM PERIOD

Station		RMSE	MAE	MAPE (%)	R^2	TIC	VAR
AREQ	GIM	7.3373	5.0974	18.2581	0.8677	0.0947	52.3991
	IRI-2016	15.0533	11.2243	32.1916	0.4431	0.2236	140.0641
	NN	5.3955	**3.3828**	**12.2011**	0.9284	0.0682	29.0872
	LSTM	5.4105	3.5515	13.8649	0.9281	0.0688	29.2576
	ANFIS-SC	6.4876	3.7198	14.4714	0.8966	0.0821	42.0815
	GBDT	**5.2713**	3.4119	13.1647	**0.9317**	**0.0666**	**27.7051**
DARW	GIM	10.1576	5.9671	24.9775	0.6382	0.1656	102.8332
	IRI-2016	11.9214	8.6817	35.4189	0.5017	0.2089	127.0952
	NN	3.4301	2.5503	12.3594	0.9587	0.0553	11.7367
	LSTM	3.8787	2.7815	13.2137	0.9473	0.0624	14.6465
	ANFIS-SC	6.0793	2.6485	11.9094	0.8704	0.0965	36.5193
	GBDT	**3.3065**	**2.4211**	**11.572**	**0.9617**	**0.0527**	**10.1007**
XMIS	GIM	12.3853	7.5434	33.1579	0.716	0.1697	153.1346
	IRI-2016	14.1275	8.8322	33.5509	0.6305	0.2135	190.2025
	NN	**4.0396**	2.8004	13.5088	**0.9698**	**0.0552**	**16.2769**
	LSTM	5.053	3.2034	14.0506	0.9527	0.0698	25.5314
	ANFIS-SC	4.4499	2.925	15.0634	0.9633	0.0607	19.7402
	GBDT	4.0719	**2.742**	**12.4267**	0.9693	0.0555	16.3566

models obtain slightly better prediction accuracy. For prediction stability, the four proposed models are generally more stable than GIM and IRI-2016 models, while the VAR values of GBDT and NN models are better than that of the other models, which shows the good prediction stability of the two models in the three IGS stations. In the magnetic storm period, the proposed GBDT still achieves about 12.7% RMSE improvement on average comparing to the rest three machine learning-based approaches.

VI. CONCLUSION

Real-time precise ionosphere TEC monitoring is critical for many aerospace applications, while the classical prediction methods based on time series analysis cannot meet the requirement. The machine learning-based approach can learn the implicit relationship between the ionosphere TEC value and the external features, which can be used to improve the precision of short-term ionosphere TEC prediction. A few machine

learning-based approaches have been proposed, while how to optimally select the machine learning approach has not been addressed, especially in severe space weather conditions.

In this article, three existing machine learning-based models, namely NN, LSTM, ANFIS, and a new model namely the GBDT, are adopted to predict the GPS VTEC time series. Three stations from the low geomagnetic latitude are selected and 4 years of data in the highest solar activity period are used to evaluate the performance of the machine learning-based models. The TEC prediction accuracy and stability are evaluated by six different indicators, and the performance of these models under extreme space weather conditions is assessed. The numerical results indicate that the ionosphere prediction accuracy in high solar activity is not as good as in the low solar activity period, but the machine learning-based approaches still significantly outperform the GIM and IRI-2016 models. The machine learning-based approach achieves about 2.9–4.7 TECU prediction accuracy during the high solar activity period and the newly introduced GBDT approach achieves the best performance out of the four models. The mean prediction accuracy improvement of GBDT subject to the rest three models is about 5.6%. The ionosphere prediction accuracy presents dependency on the season and the local time. The performance of the machine learning-based approach in the magnetic storm is also evaluated. The results indicate that the magnetic storm further deteriorates the ionosphere VTEC prediction accuracy by about 2 TECU. The magnetic storm makes the ionosphere prediction less reliable. However, the GBDT approach still achieves the best performance in the magnetic storm and the mean RMSE improvement reaches 12.7% comparing to the rest three models. In addition to the prediction accuracy, the variance of the prediction error of the four machine learning-based models presents better stability than GIM and IRI-2016. Overall, in the challenging space weather scenario, machine learning-based methods outperform the existing GIM and IRI-2016 models, and the GBDT models perform better than the other models in terms of prediction accuracy and stability under severe space weather conditions.

ACKNOWLEDGMENT

The authors would like to thank CODE for providing the GIM products. The IGS data and GIM products used in this study are acquired from the CDDIS.[1] The IRI working group provides access to the IRI model.[2] Besides, the numerical calculations in this study are performed on the supercomputing system of the Supercomputing Center of Wuhan University.

REFERENCES

[1] J. Lilensten, *Space Weather: Research Towards Applications in Europe*, vol. 344, Berlin, Germany: Springer, 2007, pp. 125–203.

[2] P. Chen, Y. Yao, and W. Yao, "Global ionosphere maps based on GNSS, satellite altimetry, radio occultation and DORIS," *GPS Solut.*, vol. 21, no. 2, pp. 639–650, 2017.

[3] C. Zhao, Y. Yuan, B. Zhang, and M. Li, "Ionosphere sensing with a low-cost, single-frequency, multi-GNSS receiver," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 881–892, Feb. 2019.

[1][Online]. Available: ftp://cddis.gsfc.nasa.gov

[2][Online]. Available: http://irimodel.org/IRI-2016/

[4] Z. Gao, M. Ge, Y. Li, Y. Pan, Q. Chen, and H. Zhang, "Modeling of multi-sensor tightly aided BDS triple-frequency precise point positioning and initial assessments," *Inf. Fusion*, vol. 55, pp. 184–198, 2020.

[5] M. Hernández-Pajares *et al.*, "The IGS VTEC maps: A reliable source of ionospheric information since 1998," *J. Geodesy*, vol. 83, no. 3–4, pp. 263–275, 2009.

[6] R. Orús, M. Hernández-Pajares, J. M. Juan, J. Sanz, and M. Garcıa-Fernández, "Performance of different TEC models to provide GPS ionospheric corrections," *J. Atmos. Sol.-Terr. Phys.*, vol. 64, no. 18, pp. 2055–2062, 2002.

[7] A. García-Rigo *et al.*, "Global prediction of the vertical total electron content of the ionosphere based on GPS data," *Radio Sci.*, vol. 46, no. 6, pp. 1–3, 2011.

[8] F. Zheng, "Research on BDS/GNSS real-time wide-area high-accuracy atmospheric delay modeling and augmentation precise point positioning applications," Ph.D. dissertation, Wuhan Univ., Wuhan, 2017.

[9] D. Bilitza *et al.*, "International reference ionosphere 2016: From ionospheric climate to real-time weather predictions," *Space Weather*, vol. 15, no. 2, pp. 418–429, 2017.

[10] C. Shi, T. Zhang, C. Wang, Z. Wang, and L. Fan, "Comparison of IRI-2016 model with IGS VTEC maps during low and high solar activity period," *Results Phys.,* vol. 12, no. 37, pp. 555–561, 2019.

[11] X. Zhang, X. Ren, F. Wu, and Q. LU, "Short-term TEC prediction of ionosphere based on ARIMA model," *Acta Geod. Cart. Sin.*, vol. 2, pp. 118–124, 2014.

[12] D. V. Ratnam, Y. Otsuka, G. Sivavaraprasad, and J. R. K. K. Dabbakuti, "Development of multivariate ionospheric TEC forecasting algorithm using linear time series model and ARMA over low-latitude GNSS station," *Adv. Sp. Res.*, vol. 63, no. 9, pp. 2848–2856, 2019.

[13] R. Tang, F. Zeng, Z. Chen, J.-S. Wang, C.-M. Huang, and Z. Wu, "The comparison of predicting storm-time ionospheric TEC by three methods: ARIMA, LSTM, and Seq2Seq," *Atmos. (Basel)*, vol. 11, no. 4, 2020, Art. no. 316.

[14] I. L. Mallika, D. V. Ratnam, Y. Ostuka, G. Sivavaraprasad, and S. Raman, "Implementation of hybrid ionospheric TEC forecasting algorithm using PCA-NN method," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 1, pp. 371–381, Jan. 2019.

[15] R. Song, X. Zhang, C. Zhou, J. Liu, and J. He, "Predicting TEC in China based on the neural networks optimized by genetic algorithm," *Adv. Sp. Res.*, vol. 62, no. 4, pp. 745–759, 2018.

[16] J. B. Habarulema, L. A. McKinnell, and B. D. L. Opperman, "A recurrent neural network approach to quantitatively studying solar wind effects on TEC derived from GPS; preliminary results," *Ann. Geophys.*, vol. 27, no. 5, pp. 2111–2125, 2009.

[17] I. Srivani, G. Siva Vara Prasad, and D. Venkata Ratnam, "A deep learning-based approach to forecast ionospheric delays for GPS signals," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1180–1184, Aug. 2019.

[18] A. Ruwali, A. J. S. Kumar, K. B. Prakash, G. Sivavaraprasad, and D. V. Ratnam, "Implementation of hybrid deep learning model (LSTM-CNN) for ionospheric TEC forecasting using GPS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1004–1008, Jun. 2021.

[19] L. Liu, S. Zou, Y. Yao, and Z. Wang, "Forecasting global ionospheric TEC using deep learning approach," *Space Weather*, vol. 18, no. 11, pp. 1–12, 2020.

[20] M. R. Ghaffari Razin and B. Voosoghi, "Ionosphere time series modeling using adaptive neuro-fuzzy inference system and principal component analysis," *GPS Solut.*, vol. 24, no. 2, pp. 1–13, 2020.

[21] A. J. Coster, E. M. Gaposchkin, and L. E. Thornton, "Real-time ionospheric monitoring system using GPS," *Navigation*, vol. 39, no. 2, pp. 191–204, 1992.

[22] S. Banville, P. Collins, W. Zhang, and R. B. Langley, "Global and regional ionospheric corrections for faster PPP convergence," *Navig. J. Inst. Navig.*, vol. 61, no. 2, pp. 115–124, 2014.

[23] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Jt. Conf. Artif. Intell.*, 1995, vol. 14, pp. 1137–1145.

[24] G. E. Hinton, "How neural networks learn from experience," *Sci. Amer.*, vol. 267, no. 3, pp. 144–151, 1992.

[25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," vol. 27, pp. 807–814, 2010.

[26] S. Haykin and N. Network, "A comprehensive foundation," *Neural Netw.*, vol. 2, 2004, Art. no. 41.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.,* vol. 9, no. 8, pp. 1735–1780, 1997.

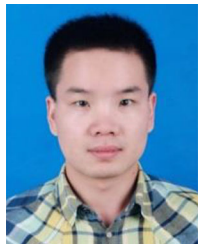
[28] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1745–1749, Oct. 2017.
[29] J.-S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man. Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.
[30] M. R. Ghaffari Razin and B. Voosoghi, "Ionosphere time series modeling using adaptive neuro-fuzzy inference system and principal component analysis," *GPS Solut.*, vol. 24, no. 2, pp. 1–13, 2020.
[31] K. Benmouiza and A. Cheknane, "Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting," *Theor. Appl. Climatol.*, vol. 137, no. 1–2, pp. 31–43, 2019.
[32] H. Sanikhani and O. Kisi, "River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches," *Water Resour. Manage.*, vol. 26, no. 6, pp. 1715–1729, 2012.
[33] Z. Wei, Y. Meng, W. Zhang, J. Peng, and L. Meng, "Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau," *Remote Sens. Environ.*, vol. 225, pp. 30–44, 2019.
[34] R. Sun *et al.*, "Improving GPS code phase positioning accuracy in urban environments using machine learning," *IEEE Internet Things J.,* vol. 8, no. 8, pp. 7065–7078, Apr. 2021.
[35] R. Sun, G. Wang, W. Zhang, L. T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," *Appl. Soft. Comput. J.*, vol. 86, 2020, Art. no. 105942.
[36] M. Alizamir, S. Kim, O. Kisi, and M. Zounemat-Kermani, "A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions," *Energy*, vol. 197, 2020, Art. no. 117239.
[37] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
[38] I. G. Richardson and H. V. Cane, "Geoeffectiveness (Dst and Kp) of interplanetary coronal mass ejections during 1995–2009 and implications for storm forecasting," *Space Weather*, vol. 9, no. 7, 2011, Art. no. 7005.
[39] K. F. Tapping, "The 10.7 cm solar radio flux (F10. 7)," *Space Weather*, vol. 11, no. 7, pp. 394–406, 2013.
[40] L. Liu and Y. Chen, "Statistical analysis of solar activity variations of total electron content derived at jet propulsion laboratory from GPS observations," *J. Geophys. Res. Sp. Phys.*, vol. 114, no. A10, 2009, Art. no. A10311.
[41] M. S. Bagiya, H. P. Joshi, K. N. Iyer, M. Aggarwal, S. Ravindran, and B. M. Pathan, "TEC variations during low solar activity period (2005-2007) near the equatorial ionospheric anomaly crest region in India," *Ann. Geophys.*, vol. 27, no. 3, pp. 1047–1057, 2009.
[42] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross-validation," *J. Mach. Learn. Res.*, vol. 5, pp. 1089–1105, 2004.
[43] J. Kong, F. Li, Y. Yao, Z. Wang, W. Peng, and Q. Zhang, "Reconstruction of 2-D/3-D ionospheric disturbances in high-latitude and arctic regions during a geomagnetic storm using GNSS carrier TEC: A case study of the 2015 great storm," *J. Geodesy*, vol. 93, no. 9, pp. 1529–1541, 2019.
[44] Y. Yao, L. Liu, J. Kong, and C. Zhai, "Analysis of the global ionospheric disturbances of the March 2015 great storm," *J. Geophys. Res. Sp. Phys.*, vol. 121, no. 12, pp. 12157–12170, 2016.
[45] D. Okoh *et al.*, "Storm-time modeling of the African regional ionospheric total electron content using artificial neural networks," *Space Weather*, vol. 18, no. 9, 2020, Art. no. e2020SW002525.

Yi Han received the M.Sc. degree in geodesy and surveying engineering from Wuhan University, Wuhan, China, in 2019. He is currently working toward the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University.

His current research interests include the applications of LEO constellation optimization and ionosphere.

Lei Wang (Member, IEEE) received the Ph.D. degree in electronical engineering and computer science (EECS) from Queensland University of Technology, Brisbane, Australia, in 2015.

He is currently an Associate Research Fellow with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. His research interests include GNSS precise positioning, LEO orbit determination, and LEO navigation augmentation and indoor positioning.



Wenju Fu received the Ph.D. degree in geodesy and surveying engineering from Chang'an University, Xi'an, China, in 2018.

He is currently a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. His research interests include GNSS precise point positioning, precise satellite clock, and LEO applications.

Haitao Zhou received the M.S. degree in geodesy and surveying engineering from Information Engineering University, Zhengzhou, China, in 2016. He is currently working toward the Ph.D. degree in geodesy and surveying engineering with Wuhan University, Hubei, China.

His research interests include surveying data processing and GNSS precise positioning.

Tao Li was born in Jiangxi, China, in 1995. He received the B.Sc. degree in surveying and mapping engineering from the China University of Mining and Technology, Xuzhou, China, in 2018. He is currently working toward the Ph.D. degree in geodesy and surveying engineering with Wuhan University, Wuhan, China.

His research interests include troposphere delay modeling, ionosphere, and the LEO navigation augmentation technology.

Ruizhi Chen received Ph.D degree in geodesy from Finish Geodetic Institute, Masala, Finland, in 1991.

He is currently a Professor and the Director with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was an Endowed Chair Professor with Texas A&M University-Corpus Christi, Corpus Christi, TX, USA, the Head and a Professor with the Department of Navigation and Positioning, Finnish Geodetic Institute, Masala, Finland, and the Engineering Manager with Nokia, Espoo, Finland. He has authored or coauthored two books and more than 200 scientific papers. His current research interests include indoor positioning, satellite navigation, and location-based services.