A Stacked Machine Learning Model for the vertical Total Electron Content Forecasting

Eric N. Asamoah[(1)(2)]*, Massimo Cafaro[(1)], Italo Epicoco[(1)], Giorgiana De Franceschi[(2)] and Claudio Cesaroni[(2)]

(1) University of Salento, Department of Engineering for Innovation, Lecce, Italy; e-mail: ericnana.asamoah@unisalento.it;
massimo.cafaro@unisalento.it; italo.epicoco@unisalento.it

(2) INGV, Environment Department, Rome, Italy; e-mail: giorgiana.defranceschi@ingv.it; claudio.cesaroni@ingv.it

Abstract

This study focuses on the application of an advance machine learning model (stacked) for forecasting the ionospheric vertical Total Electron Content (vTEC) from 1 to 24 hours in advance. Data are provided by the Global Navigation Satellite Systems (GNSS) receiver installed at Noto, Italy (NOT1 (lat:36.53, 14.59) data). The stacked model performance is compared with ANN, CNN and ELM. The time series used include vTEC and external drivers (solar and geomagnetic indices) which span from January 1, 2011 to December 31, 2011. The dataset is partition into a training (80%), validation (10%) and test (10%). Proper external drivers have been selected by ranking their importance of each driver in relation to the vTEC dynamics at different hours of the day. The most important external drivers resulted to be the 10.7 cm solar flux (F10.7) and the Geomagnetic Auroral Electrojet index (AE), used together with the vTEC times series as the input for our model. To measure the performance of the models, we make use of the statistical parameters such as root mean square error (RMSE) and coefficient of determination (R^2). In the period analyzed, the stacked model showed the best performance on the test dataset with $R^2 = 0.97$ and $RMSE = 0.16$ TECU, followed by ANN, CNN, and ELM with $R^2 = 0.96$ and $RMSE = 0.17$ TECU, 0.18 TECU and 0.18 TECU respectively. In conclusion, the stacked model is promising irrespective of its challenges.

1 Introduction

The ionosphere of the Earth is a partly ionized gas that surrounds the planet from about 50 km up to 1000 km and above. This is a complex system that is disrupted by various factors including geomagnetic storms, Coronal Mass Ejection (CME) and other space weather phenomenon and forcing from lower atmosphere [1]. The ionosphere changes throughout the day, and all of the elements of this highly coupled system undergo several timeframe changes, from impulsive solar flares or auroral intensifications (which lasts for around minutes) to solar cycle lengths (≈ 11 years). Global Navigation Satellite Systems (GNSS) applications such as navigation, positioning and timing, rely on L-band signals passing through the ionosphere received at ground. The integrity, accuracy and availability of GNSS is greatly influence by space weather conditions and in particular the ionosphere is the largest contributor to the error budget for the GNSS positioning applications. On other hands, the measurements of GNSS provides useful information about the distribution of electrons in the ionosphere allowing measuring the integrated umber of free electron along the signal path from satellite to ground i.e. the slant TEC (see e.g. [2]) that can be projected to vertical obtaining the so called vertical Total Electron Content (vTEC). In order to understand and compare the observed values of vTEC, many models have been developed. These models may be divided into two types in general: (i) Traditional methods; (ii) Machine Learning based methods and recently many of them are based on Machine Learning techniques. Machine Learning models are widely used, and flourishing in several fields of study, including image processing and computer vision for classification, detection, recognition, language translation and regression problems. In Gebreah K. Z. et al, 2021 [3], the authors presented a data-driven forecasting of ionospheric vTEC using a Long-Short Time Memory (LSTM) deep recurrent neural network. In the process of selecting the input parameters to train the algorithm, they used the Random Forest algorithm to perform regression analysis and estimate the importance of input parameters. Relative importance of 34 different parameters including the solar flux, solar wind density, and speed of the three components of interplanetary magnetic field, Lyman-alpha, the Kp, Dst and Polar Cap (PC) indices were analyzed. The LSTM method was applied to forecast the vTEC up to 5 hours ahead. A good forecast was achieved with low RMSE but the RMSE increases as they forecast further into the future.

In this work, the goal is to develop a new Machine Learning based algorithm to forecast ionosphere vTEC values by stacking several neural networks together, building one robust ensemble model. The base neural network models selected include ANN, CNN and ELM. The rest of this paper is organized as follows. In section 2, we discuss the data used and briefly recall how ANN, CNN, ELM and a stacked model work. Moreover, we also discuss how we selected the most relevant features (input parameters) for the prediction task. In Section 3 we provide and discuss extensive experimental results. Finally, in Section 4 we draw our conclusions and highlight further possible future investigation.

2 Data and Methodology

2.1 Data Preprocessing

The GNSS data is obtained from NASA's Archive of Space Geodesy (https://cddis.nasa.gov/archive/gnss/data/daily). In particular, these are 30-second GNSS data from IGS regional data collections centers which are compressed in RINEX format to the CDDIS on a daily basis. In order to extract vTEC of mid-latitude station NOT1 (lat: 36.53, lon: 14.59) from the RINEX file format, we made use of the calibration algorithm developed by Ciraolo et al, 2007 and detailed in Cesaroni et al, 2015 and Cesaroni et al., 2021. This algorithm is able to estimate the Differential Code Biases (DCBs) affecting the sTEC estimation. From the calibrated sTEC from each satellites in view, the algorithms is then able to project the sTEC into the vTEC by applying a mapping function [5, 2, 6]. To train the model, vTEC data was divided into three partitions of ratios 0.8, 0.1 and 0.1 respectively for training, validation and test, as shown in Figure 1. To understand which external drivers are suitable for our model, we ranked the external drivers in order of importance with regard to the effect they have on vTEC values. The external drivers for the year 2011 are from https://cdaweb.gsfc.nasa.gov and reported in table 1.

Indices	Parameters
Aurora Electrojet	AU, AL, AE
Geomagnetic	Dst, SYM/H, Kp
Magnetic and Solar	F10.7, SSN,
	proton density, flow pressure,
	velocity (Vx, Vy , Vz)(km/s),
	solar wind (Proton QI)
	magnitude of average field vector (nT),
	and (Bx, By, Bz)(GSE)

Table 1. External drivers used for our study.
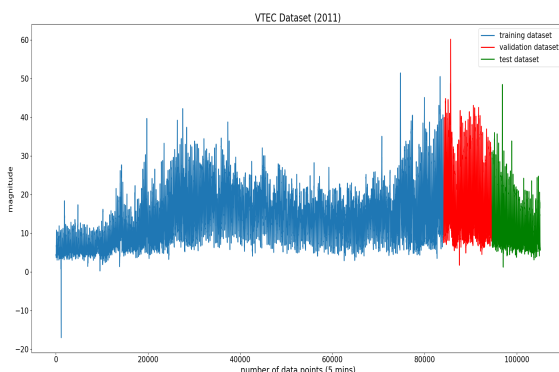
Figure 1. This figure shows the vTEC dataset (NOT1; lat: 36.53, lon: 14.59) which has been been down-sampled to 5 minutes intervals. The data in blue color are used for training the neural network models, the ones in red are for validation and the green ones are used for testing the model.
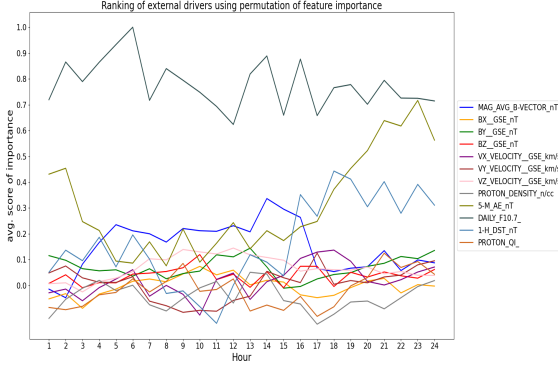
2.2 Feature Ranking

Machine Learning has been extensively exploited in different fields in order to help and support decision making, but it is important to understand why a particular output is produced according to the input (features) used to train the models. For this reason, we use a feature ranking technique called permutation of feature importance. To determine the importance of the external drivers, we used 12 Machine Learning models including: Random Forest, Support Vector Regressor (SVR), Extreme Gradient Boosting, Gradient Boosting, linear regression, LASSO, Adaptive Boosting, Decision Tree, Extra Tree, KNN, bagging (using SVR as estimator), voting (using as estimators: Adaptive Boosting, Decision Tree, KNN, Random Forest, Extra Tree, Gradient Boosting, Extreme Gradient Boosting). Ensembling these models together ensures robustness by reducing the variance, since there is no specific Machine Learning model which is always optimal (no free lunch theorem for Machine Learning) [7]. Before using the permutation importance technique, we first checked for multicollinearity between the external drivers and removed the correlated ones by using a threshold of 0.5 (in particular sym/h correlated with Dst (0.95) & AL (0.50), SSN correlated with F10.7 (0.89), flow pressure correlated with proton density (0.81), AL correlated with AU (0.5) & Dst (0.5), kp correlated with AE (0.71), flow pressure (0.5), & mag_avg (0.52) and AU correlated with AE (0.86) & kp (0.69)). Since the lag between the solar and geomagnetic forcing and the response of the ionosphere is not fully known understood, we performed 24 different feature ranking at different hours by using the external drivers as input variable and vTEC values as output at different hours with regard to the hour of interest. The overall ranking of the features (external drivers) are depicted in Figure 2. The best two features have been selected for the predictive model at different hours. The most important ones include F10.7, Dst, AE and mag_avg_B_vector.

2.3 Predictive Model

In order to build our predictive model, we required input dataset including the two best-ranked external features for different hours with 24 hours and vTEC times series. We took the previous 24 hours of the vTEC and the two best external drivers times series as the features of each instant of the training dataset. In this section we briefly introduce the different ML algorithms used in this work.

2.3.1 Artificial Neural Network (ANN)

Neural networks [8] are learning machines potentially containing a large number of neurons, which are connected in a layered fashion. Learning is achieved by adjusting the synaptic weights to minimize a predefined cost function. The back-propagation algorithm was a breakthrough, since it enabled training neural networks on a set of input-output

Figure 2. This figure shows the ranking of external drivers. F10.7 was ranked first throughout the 24 hours, the second ranked features alternated among AE, Dst, and Mag_avg. AE ranked second for 1st-4th hours, Mag_avg was second for 5th-11th hours and again AE ranked second on the 12th hour, then Mag_avg was selected for 13th-15th hours. Dst took over the second position for 16th-18th hours and finally AE was selected for 19th-24th hours.

training samples easily (even though training is computationally intensive) [9]

2.3.2 Convolutional Neural Network (CNN)

Different models have been created to match the characteristics of the data for various applications [10]. A breakthrough known as Convolutional Neural Network (CNN) appeared in the late 1980s, in order to analyze images. Indeed, in the original model the neurons were connected to resemble the organization of the visual cortex of a cat. In our architecture we have three small CNNs whose output are fused and passed to a fully connected network.

$$\mathcal{O}(i,j) = \sum_{t=1}^{m} \sum_{r=1}^{m} h(t,r)I(i+t-1, j+r-1) \quad (1)$$

where $m < l$.

2.3.3 Extreme Learning Machine

Extreme Learning Machine (ELM) [11] is a feedforward neural network with one or more layers of hidden nodes, and it requires tuning for hidden node parameters as well as the weights that connect inputs to hidden nodes. The weights of the nodes of hidden layers may be randomly chosen and never updated, or they may be passed down unchanged from predecessors to successors neurons. The output weights of hidden nodes are typically learnt in a single step, which is equivalent to learning a linear model in most situations. While achieving greater generalization performance, the learning speed can be thousands of times quicker than that of conventional feedforward network learning algorithms like the back-propagation (BP) method.

2.4 Stacked Model

In the quest of improving the performance of a predictive model, an ensemble approach called stacking, is based on hierarchy of the models. The stacking approach for combining other predictors was developed by Wolpert in 1992 [12]. The stacked model we developed exploits ANN, CNN, and ELM as its base learners.

3 Results

In order to evaluate the performance of the four models built (ANN, CNN, ELM and the stacked model), we plot for each model the coefficient of determination and the RMSE obtained using the test data as shown in Figure 1.

## 3.1 Results of the predictive performance on the test dataset

In Figure 3, the blue curve represents the coefficient of determination (R^2) of the stacked model, which is the best among all of the models. The model predicted 24 hourly points and the maximum of R^2 was obtained, as expected for the prediction of the 1st hour ($R^2 = 0.97$) whilst the minimum R^2 occurred on the prediction of the 13th hour ($R^2 = 0.88$) using the previous 24 hours of both vTEC and the external drivers. The average performance of the stacked model is given by $R^2 = 0.9$. The corresponding RMSE for the stacked model has a minimum of 0.16 TECU and maximum of 0.30 TECU, with an average of 0.28 TECU as shown in Figure 4. The second best of the models is the ANN, with maximum is $R^2 = 0.96$ whilst the minimum is $R^2 = 0.81$. The green curve in Figure 3 shows the coefficient of determination of ANN, the maximum R^2 occurred at the 1st hour prediction and the minimum R^2 occurred at the 8th hour. Their corresponding RMSE are 0.17 and 0.33 TECU respectively as depicted in Figure 4. The average RMSE for ANN was 0.29 TECU. CNN and ELM were the least performing models with regard to both R^2 and RMSE. The maximum R^2 for both CNN and ELM is 0.96 and the minimum R^2 are respectively 0.81 for CNN and 0.71 for ELM. The RMSE for the corresponding maximum R^2 are 0.18 for both CNN and ELM, whilst regarding the minimum it is 0.39 TECU for CNN and 0.40 TECU for ELM. The average R^2 is respectively 0.86 and 0.84, and their corresponding average RMSE is respectively 0.33 TECU and 0.35 TECU. The maximum R^2 for both CNN and ELM occurred at the 1st hour prediction whilst the minimum R^2 occurred at 23rd hour for CNN and 14th-15th hour for ELM.

4 Conclusions and Future Work

In this paper, we evaluated the performance of four predictive models. The performance of our stacked model was better than that of the ANN, CNN and ELM models. It was observed that the inference speed of the models is really fast, especially for the ELM model. The training time
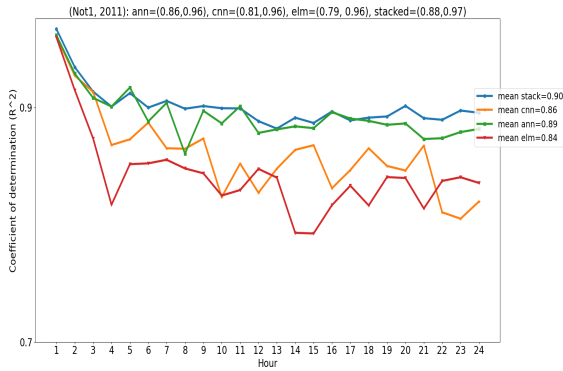
Figure 3. Coefficient of determination of each hourly prediction within 24 hours for ANN, CNN, ELM and stacked model.
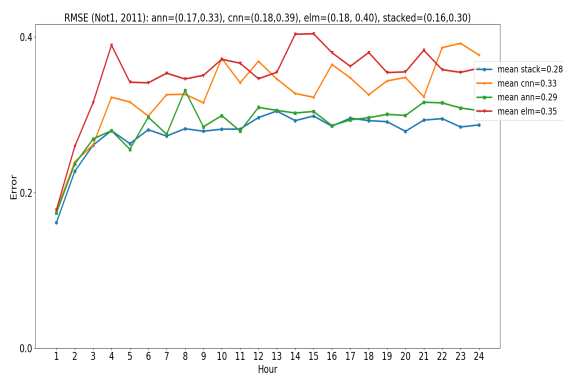
Figure 4. RMSE of each hourly prediction within 24 hours for ANN, CNN, ELM and stacked model.

of ELM is about 8-12 times faster than the training time of ANN, CNN and stacked models. The performance of the model decreases when the time lapse increases. Future work includes training our models on a solar cycle, ie., ≈ 11 years, and developing even better ensemble models for prediction. This approach can be applied in order to support global, regional and local forecasting systems that are very important for providing a reliable space weather service to many applications based on the use of GNSS technology.

References

[1] M. Mendillo, "Storms in the ionosphere: Patterns and processes for total electron content," *Reviews of Geophysics*, vol. 44, no. 4, 2006.

[2] C. Cesaroni, L. Spogli, L. Alfonsi, G. De Franceschi, L. Ciraolo, J. F. G. Monico, C. Scotto, V. Romano, M. Aquino, and B. Bougard, "L-band scintillations and calibrated total electron content gradients over brazil during the last solar maximum," *Journal of Space Weather and Space Climate*, vol. 5, p. A36, 2015.

[3] G. K. Zewdie, C. Valladares, M. B. Cohen, D. J. Lary, D. Ramani, and G. M. Tsidu, "Data-driven forecasting of low-latitude ionospheric total electron content using the random forest and lstm machine learning methods," *Space Weather*, vol. 19, no. 6, p. e2020SW002639, 2021.

[4] C. Cesaroni, L. Spogli, A. Aragon-Angel, M. Fiocca, V. Dear, G. De Franceschi, and V. Romano, "Neural network based model for global total electron content forecasting," *Journal of Space Weather and Space Climate*, vol. 10, p. 11, 2020.

[5] L. Ciraolo, F. Azpilicueta, C. Brunini, A. Meza, and S. M. Radicella, "Calibration errors on experimental slant total electron content (tec) determined with gps," *Journal of Geodesy*, vol. 81, no. 2, pp. 111–120, 2007.

[6] C. Cesaroni, L. Spogli, and G. De Franceschi, "Ionoring: Real-time monitoring of the total electron content over italy," *Remote Sensing*, vol. 13, no. 16, p. 3290, 2021.

[7] S. Bobek, P. Bałaga, and G. J. Nalepa, "Towards model-agnostic ensemble explanations," in *International Conference on Computational Science*. Springer, 2021, pp. 39–51.

[8] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[10] S. Theodoridis and K. Koutroumbas IV, "Clustering algorithms iv," *Pattern Recognition (Fourth Edition). Academic Press, Boston*, pp. 765–862, 2009.

[11] A. Akusok, K.-M. Björk, Y. Miche, and A. Lendasse, "High-performance extreme learning machines: a complete toolbox for big data applications," *IEEE Access*, vol. 3, pp. 1011–1025, 2015.

[12] D. H. Wolpert, "Stacked generalization," *Neural networks*, vol. 5, no. 2, pp. 241–259, 1992.