Neural Network Models for Ionospheric Electron Density Prediction: A Neural Architecture Search Study

Yang Pan[1], Mingwu Jin[1], Shunrong Zhang[2], Simon Wing[3], and Yue Deng[1]

[1] University of Texas at Arlington, Arlington, TX, USA.

[2] Haystack Observatory, Massachusetts Institute of Technology, Westford, MA, USA

[3] Applied Physics Laboratory, The Johns Hopkins University, Laurel, MD, USA

Corresponding author: Mingwu Jin (mingwu@uta.edu) and Yue Deng (yuedeng@uta.edu)

Key Points:

- Neural architecture search (NAS) is used to automatically find the best network structure and hyperparameters for neural network (NN) models on incoherent scatter radar (ISR) electron density data.
- A total of 16-year of data from Millstone Hill ISR are used for single-layer NNs (SLNNs), deep NNs (DNNs) and their NAS counterparts.
- NN models can reveal more finer details of electron density patterns than the empirical ionospheric model and NAS models can improve over manually tuned NN models, but the improvement is limited. The limited improvement could be due to the network complexity and the limitation of fully connected NN without the time histories of input parameters.

Abstract

Specification and forecast of ionospheric parameters, such as ionospheric electron density (*Ne*), have been an important topic in space weather and ionosphere research. Neural networks (NNs) emerge as a powerful modeling tool for *Ne* prediction. However, heavy manual adjustments are time consuming to determine the optimal NN structures. In this work, we propose to use neural architecture search (NAS), an automatic machine learning method, to mitigate this problem. NAS aims to find the optimal network structure through the alternate optimization of the hyperparameters and the corresponding network parameters. A total of 16-year data from Millstone Hill incoherent scatter radar (ISR) are used for the NN models. One single-layer NN (SLNN) model and one deep NN (DNN) model are trained with NAS, namely SLNN-NAS and DNN-NAS, for *Ne* prediction and compared with their manually tuned counterparts based on previous studies, denoted as SLNN and DNN. Our results show that SLNN-NAS and DNN-NAS outperformed SLNN and DNN, respectively. These NN models can reveal more finer details of *Ne* patterns than the empirical ionospheric model developed using traditional data fitting approaches. DNN-NAS yields the best prediction accuracy measured by quantitative metrics and rankings of daily pattern prediction. The limited improvement of NAS is likely due to the network complexity and the limitation of fully connected NN without the time histories of input parameters.

Plain Language Summary

Neural network (NN) models have garnered significant attention for their application in predicting physical parameters in the ionosphere, notably ionospheric electron density (*Ne*). In this study, we introduce a novel approach aimed at enhancing the performance of NN models by employing the advanced technique known as neural architecture search (NAS). Leveraging a dataset spanning sixteen years of *Ne* measurements obtained from the incoherent scatter radar located at the Millstone Hill observatory, we conduct a comprehensive analysis. This analysis encompasses training both manually calibrated NN models and NN models optimized via NAS. The NN models fine-tuned through NAS achieve a notable improvement in their ability to predict *Ne* when compared to their manually adjusted counterparts. This improvement underscores the efficacy of NAS in optimizing neural network hyperparameters for ionospheric modeling. Furthermore, we delve into a thorough exploration of the factors contributing to the somewhat limited improvements observed in the context of our current dataset. This investigation yields valuable insights and prompts valuable discussions on the potential avenues for further refinement in ionospheric prediction methodologies.

1 Introduction

The incoherent scatter radar (ISR) can provide direct measurements of ionospheric parameters, such as electron density (*Ne*), plasma temperature, and line of sight ion velocity. The altitudinal (range) variation of these parameters is measured continuously over time by the ISR. However, most ISRs operate for campaign purposes but not on a daily basis. Figure 1 shows an example of *Ne* around 350 km at the Millstone Hill station in 2012, where a lot of data are missing. Therefore, a model that can fill the observational data gaps for these parameters under real solar/geomagnetic conditions would be desired for various space weather and ionospheric research purposes.

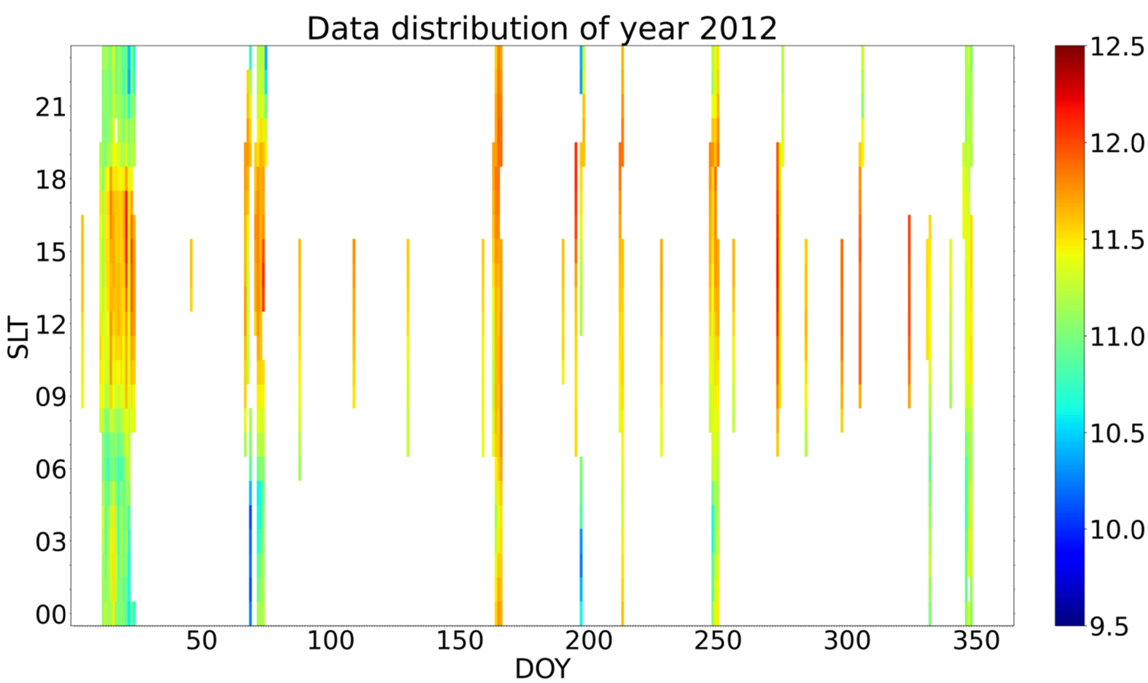


Figure 1 *The ISR records of Ne in the logarithmic scale around 350 km altitude in 2012. Horizontal axis: day of year (DOY); vertical axis: solar local time (SLT); the intensity represents logarithmic electron density ($log_{10}Ne$), while the blank space represents missing records. Most of the region is in blank, indicating the irregularity of ISR's operation.*

Conventionally, the empirical models were developed to provide this information. For example, a global model, international reference ionosphere (IRI) *Bilitza* [2001] and IRI-2016

[*Bilitza et al.*, 2017], takes primarily ionosonde observations to generate 3D distributions of ionospheric parameters. The ISR ionospheric model (ISRIM) [*Holt et al.*, 2002] has been built for multiple ISRs around the world developed initially for Millstone Hill ISR observations in the time and vertical domains [*Holt et al.*, 2002]. Additional regional models beyond local vertical variations were also developed near Millstone Hill as well as in the North America longitudes. These statistical models took a binning and fitting approach to construct an empirical model in space and time [*Zhang and Holt*, 2007]. In each bin, the sequential least-squares fit is based on the normalized F10.7 and Ap3 indices, especially with the new introduced parameter F10.7p [*Liu et al.*, 2006; *Richards et al.*, 1994] for better linear fitting [*Zhang and Holt*, 2007]. However, ISRIM was designed to provide ionospheric climatology where altitudinal and temporal variations are represented by smooth analytical models. The artificial neural network (ANN) models may be trained to better fill the data gaps or to predict these parameters.

The neural network regression models have been developed for space weather research (see for example [*S Wing et al.*, 2005]). A single hidden layer ANN with 18 neurons was used to derive ionospheric models in order to evaluate the long-term trends of *Ne* for the DMSP data [*Y Cai et al.*, 2019; *Yue et al.*, 2018]. The deep neural network (DNN) was used to model *Ne* to reconstruct the dynamics in the plasmasphere [*Bortnik et al.*, 2016]. To offer the short-term variations, a three-dimensional dynamic electron density (DEN3D) model [*XN Chu et al.*, 2017; *X Chu et al.*, 2017] is also developed for plasmasphere using DNN with enhanced number of drivers of F10.7 and AL apart from SYM-H. Several global ANN models have been proposed to predict ionospheric *Ne*. The ANN-based ionospheric models (ANNIM-2D and ANNIM-3D) have been proposed using a single-layer NN (SLNN) and more than 10-year data from the GPS-RO missions [*Gowtam et al.*, 2019; *Sai Gowtam and Tulasi Ram*, 2017; *Tulasi Ram et al.*, 2018]

(CHAMP, GRACE, and COSMIC) and the ground-based Digisonde GIRO (with 864 spatial grids for ANNIM-3D). Another global model (with 864 sub-models) was also proposed using COSMIC data [*Habarulema et al.*, 2021], where each sub-model adapted a SLNN. A three-hidden-layer DNN was used for a global 3D model ("ANN-TDD") based on COSMIC, Fengyun-3C and Digisonde data [*Li et al.*, 2021]. The most recent work combined DNN with IRI ("ANN-IRI") to improve *Ne* prediction compared to pure data-driven ANNs, particular in the lower ionosphere [*Yang and Fang*, 2023]. These pioneer models reproduce the large-scale ionospheric phenomena and generally outperform the monthly-average model of IRI-2016 during the quiet time. However, firstly, the radio occultation (RO) measured *Ne* assumes the spherical symmetry which is the major source of errors when retrieving from vertical profiles [*Lei et al.*, 2007]. Secondly, the aforementioned NN models usually have a worse prediction performance during the storm time than IRI-2016 with the STORM option on (specifically tailored for predictions during the storm time). One reason is that the storm events are comparatively taking up a smaller percentage in all the data used for the model training (i.e. not focusing on storm time behaviors), thus leading to inferior *Ne* prediction of these NN models during the storm time. Furthermore, these NN models usually chose the network structures and hyperparameters manually. Not only is the manual tuning tedious (e.g. thousands of experiments were used to find a good 3-hidden-layer network structure [*Li et al.*, 2021]), but also these models could only achieve sub-optimal prediction performance.

To address this issue with NN models for *Ne* prediction, we use an automatic optimization algorithm, so called neural architecture search (NAS) to optimize a single hidden layer NN (SLNN) and a deep NN (DNN) model and compare their performance. As our goal is to introduce NAS for optimization of NN models of *Ne* prediction, we used Millstone Hill ISR data at a fixed altitude (~350 km) from 2003 to 2018 since the data around this altitude are abundant and likely relevant to the low-earth-orbit (LEO) missions, such as CHAMP and the upcoming Geospace Dynamics Constellation (GDC) mission. In Section 2, we introduce neural network and NAS for network structure and hyperparameter optimization. Then we describe the Millstone Hill ISR data and experiments in Section 3. The summary results and cases study results are presented in Section 4. The discussion and conclusion are given in Sections 5 and 6, respectively.

2 Methodology

2.1 Neural networks (NNs)

Neural network (NN) is one of the most powerful machine learning methods for regression and classification. Usually, the neural network consists of the input layer, the hidden layer(s), and the output layer. Each hidden layer is made of multiple nodes, so called neurons. Each neuron performs a non-linear activation of the weighted sum of outputs from the previous layer. When the number of the hidden layers is equal to or greater than two, the NN is called the deep neural network (DNN) otherwise the single-layer neural network (SLNN). Given the input and output variables $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, a DNN model makes prediction as $\boldsymbol{y} = f(\Theta, \boldsymbol{x}|\Lambda)$, where Θ is the trainable parameters (i.e. weights and biases connecting neurons) and Λ is the hyperparameters defining the network structure and training conditions (such as the number of layers, the number of neurons in each layer, dropout, optimizer, learning rate, etc.). If Λ is fixed and the training data are X^{train} and Y^{train}, Θ can be optimized by the following training:

$$\Theta^{*} = arg\,\min_{\Theta} loss\left(y^{train}, f\left(\Theta, x^{train} \middle| \Lambda\right)\right), for\left(x^{train}, y^{train}\right) \in \{X^{train}, Y^{train}\}\#(1)$$

where "*loss*" is the loss function measuring the overall difference between the observations and the model predictions on the training data.

However, Equation (1) only optimizes on Θ for a fixed network, i.e., fixed Λ. Based on the task and data, the performance of DNN is also dependent on the hyperparameters Λ. Manually tuning these hyperparameters could become tedious and time consuming and lead to unsatisfactory results. The search algorithms were developed to obtain the optimal solution automatically in a pre-defined hyperparameter space as described in the next section.

2.2 Neural Architecture Search (NAS) through AutoKeras

Automatic machine learning (AutoML) has become a viral research topic as machine learning is widely applicable in many fields [*Hutter et al.*, 2019]. It enables researchers in the field other than machine learning to build their models more efficiently. Neural architecture search (NAS) [*Elsken et al.*, 2019] is one subject of AutoML and aims to search to the best NN for a given task and dataset, whose flow chart is summarized in Figure 2. Reinforcement learning [*Baker et al.*, 2016; *Zoph and Le*, 2016] was first proposed for NAS, followed by gradient methods [*H Cai et al.*, 2018a; *Luo et al.*, 2018], evolutionary algorithms [*Desell*, 2017; *Guo et al.*, 2020; *Real et al.*, 2017; *Suganuma et al.*, 2017], and network morphism [*H Cai et al.*, 2018b; *Elsken et al.*, 2017; *Jin et al.*, 2019]. NAS aims to find the optimal network structure through the following alternative optimization,

$$\Lambda^{*} = arg\,\min_{\Lambda} cost\left(\boldsymbol{y}^{val}, f(\Theta^{*}, \boldsymbol{x}^{val}|\Lambda)\right), for\ (\boldsymbol{x}^{val}, \boldsymbol{y}^{val}) \in \{X^{val}, Y^{val}\}\#(2)$$

$$\Theta^* = arg\min_{\Theta} loss\left(\boldsymbol{y}^{train}, f\left(\Theta, \boldsymbol{x}^{train} \middle| \Lambda^*\right)\right), \text{for}\left(\boldsymbol{x}^{train}, \boldsymbol{y}^{train}\right) \in \{X^{train}, Y^{train}\}\#(3)$$

where the data are divided into the training set $\{X^{train}, Y^{train}\}$ and the validation set $\{X^{val}, Y^{val}\}$. While "*cost*" is the cost function measuring the model prediction error on the validation data $\{X^{val}, Y^{val}\}$, and "*loss*" is the loss function measuring the model fitting error on the training data $\{X^{train}, Y^{train}\}$ with a fixed Λ^*.

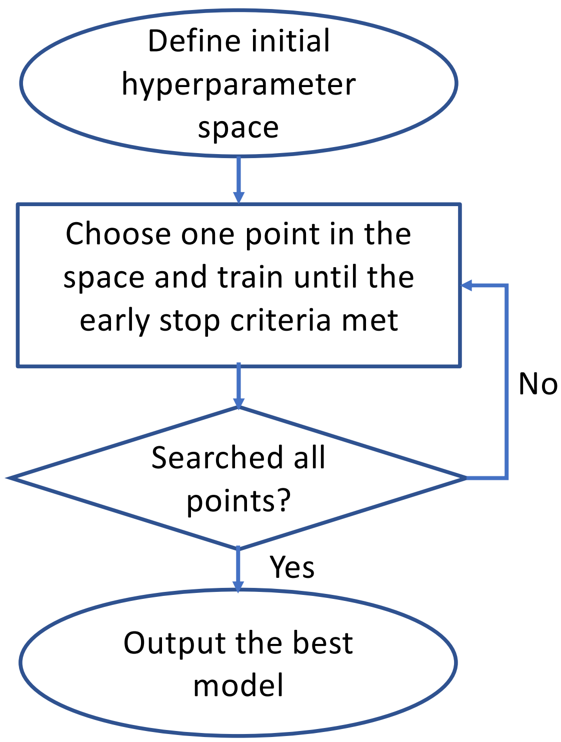


Figure 2 *Flow chart of Neural Architecture Search (NAS).*

AutoKeras [*Jin et al.*, 2019] with a high-level user interface is a NAS method based on network morphism, which modifies the NN using the morphism operations, such as inserting a layer or adding a skip-connection. To search the optimal network structure, a hierarchical tree structure is used, whose basic component is the node. For instance, the mother node is an abstract idea of the NN configuration, which is followed by a child node consisting of dense layers, activation layers, normalization layers, etc. The other child nodes include learning rate and training optimizer.

Each child node can serve as the parent node for the nodes connected at the next level, and a tree structure is conducted. Finally, the leaf is an end node without any child node. The hyperparameter space defined in Table 1 is the result of large number of empirical searches with different combinations. Neuron number no greater than 64 has already offered decent result for both SLNN and DNN. For DNN, the layer number is refrained to no more than 4 based on the literatures and our preliminary trials. The most noticeable is the learning rate search polls. To achieve stable and converging models, larger learning rates fit the SLNNs while DNNs prefers comparatively smaller ones. The reason is that a more complicated neural network structure requires more fine tuning, and hence a smaller learning rate will have a higher chance of leading to a more stable model as judged by the loss curves. However, a lower learning rate does not guarantee a smaller converged loss value. Thus, manual tuning on learning rates becomes undesirable with consideration on the efficiency. Besides, Adam optimizer [*Kingma and Ba*, 2014] is fixed as the training optimizer for all the models which is not explicitly mentioned in the table.

Table 1 *Hyperparameter space of NAS. The candidates in each hyperparameter poll are the optimal results of multiple trials. For instance, the single layered architecture prefers a larger learning rate than the deep neural architecture.*

Hyperparameter	**Range**
Number of layers	SLNN: [1] DNN: [2, 3, 4]
Neuron number	[16, 18, 20, …, 64]
Learning rate	SLNN: 9e-04, 8e-04, …, 1e-04
	DNN: 5e-04, 4e-04, …, 5e-05

Three representative search algorithms in AutoKeras for NAS are: random search, greedy search, and Bayesian optimization. A trial is defined as a round of optimization of Equation. (2) with a single set of hyperparameter configuration when the early stopping criterion, i.e., no significant improvement of the objective function, is met. Besides, the maximum allowed number of trials is defined at the beginning. For those three search algorithms, random search randomly picks a hyperparameter configuration without repetition for each trial until the number of trials is reached. Apparently, the random search suffers the inefficiency. The greedy search selects a node with a probability inversely proportional to the number of leaves of that node. The other hyperparameters in the search space will be picked randomly first, then as the previous best trial to form a trial configuration. Therefore, the advantage for the greedy search over the random search is that the search can always return to the best trial when the new configuration does not offer better performance. Each trail of the Bayesian optimization (BO) consists of a loop of update, generation, and observation. A neural network kernel function is defined to measure the edit-distance between two network structures, which will enable the Gaussian process-based update of the network architecture. Upper-confidence bound is used for the cost function, whose optimization leads to generation of the next network architecture Λ^*. The observation is to obtain the optimal weights Θ^* for the new network architecture as shown in Equation. (3). These three steps repeat until the pre-defined trial number is reached. More details of AutoKeras can be found in [*Jin et al.*, 2019]. During the trials, we found that the greedy algorithm had the advantage over the remaining search algorithms. Thereafter, the greedy algorithm is fixed for all the following experiments.

In this work, we developed several models for Millstone Hill *Ne* prediction: 1) single-layer neural network with an arbitrary structure (SLNN) (18 neurons in the hidden layer [*Y Cai et* *al.*, 2019; *Yue et al.*, 2018]); 2) SLNN with NAS (SLNN-NAS); 3) deep neural network with an arbitrary structure (DNN) (three hidden layers with 24, 22, and 20 neurons, respectively [*Li et al.*, 2021]); and 4) DNN with NAS (DNN-NAS).

3 Data and experiments

The Millstone ISR *Ne* data at the fixed altitude of ~350 km from 2003-2018 were used for training and test of different NN prediction models. The input variables are year, day number of year (DOY), solar local time (SLT, hour), daily F10.7 index (solar flux unit or sfu), and 3-hourly Ap index (Ap3), in which the cyclic sine and cosine are applied on DOY (DOY_s and DOY_c in equation. (4)) and SLT (SLT_s and SLT_c in equation. (5)) to reflect the periodic changes of these two input variables as suggested by previous studies [*Athieno et al.*, 2017; *Habarulema et al.*, 2021] as well as more stable training. If not specifically elaborated, the output variable *Ne* stands for the logarithmic electron density (i.e. *Ne* is equivalent to $\log_{10}Ne$, particularly for the numerical values) in the following sections.

$$DOY_s = \left(\sin 2\pi \times \frac{DOY}{365} + 1\right)/2, DOY_c = \left(cos\, 2\pi \times \frac{DOY}{365} + 1\right)/2 \quad (4)$$

$$SLT_s = \left(\sin 2\pi \times \frac{SLT}{24} + 1\right)/2, SLT_c = \left(cos\, 2\pi \times \frac{SLT}{24} + 1\right)/2 \quad (5)$$

Table 2 *Data setting and the conditions to clean ISR data. The ISR data has the greatest number of observations near height of 350km, which indicates the data availability is of our major consideration. The filters on two F10.7 and Ap3 would rule out high intensity geophysical events.*

<table>
<tr><th>Parameter</th><th colspan="2">Values</th></tr>
<tr><td rowspan="3">Years</td><td>Training</td><td>2003 to 2018 except the val&test sets</td></tr>
<tr><td>Validation</td><td>[2010, 2015]</td></tr>
<tr><td>Test</td><td>[2007, 2012]</td></tr>
<tr><td>F10.7</td><td colspan="2">≤ 300 sfu</td></tr>
<tr><td>Ap3</td><td colspan="2">≤ 80</td></tr>
<tr><td>Altitude</td><td colspan="2">~350 km</td></tr>
<tr><td>$\boldsymbol{Ne}$</td><td colspan="2">$[\log_{10}(5 \times 10^9), \log_{10}(3 \times 10^{12})]$ el/m^3</td></tr>
</table>

A total of 16 years of ISR data from 2003 to 2018 were used. Year 2010 and 2015 were selected as validation set, while year 2007 and 2012 were reserved as test set. Remaining 12 years of data were used for training. We first cleaned the ISR data following the conditions in Table 2. Specifically, the data corresponding to high solar activity and intense earth magnetic conditions (with F10.7 over 300 sfu and Ap3 greater than 80 units), which take about only 2% of whole dataset, were discarded following the previous work [*Y Cai et al.*, 2019]. The *Ne* values were also confined to the range of $[5 \times 10^9, 3 \times 10^{12}]$ el/m^3. Furthermore, the noisy data that show isolated peaks/troughs or irregular time intervals in daily patterns were discarded. Finally, the remaining data were binned to a one-hour interval. One hour cadence was chosen to balance short-term variability in data and temporal resolution of the model. We also assured that the training, validation, and test sets followed the similar distribution of that of the overall *Ne*. After all these preprocessing of data, the training/validation/test set include 8,052/1,461/1,970 data records, respectively.

We used the mean absolute error (MAE), root mean squared error (RMSE), and relative error (RE) of the test data as the quantitative measures for the prediction performance. The Bland-Altman plots were used to interrogate the agreement between model output and ground truth *Ne*. We also quantitatively compared the predicted annual and day-to-day variations for all models supplemented by rankings of a daily variation prediction.

4 Results

In this section, the best network structure for the NAS models and the search for the best learning rates for all the models are presented first. Then, the prediction performance is evaluated statistically using MAE, RMSE, RE, and Bland-Altman plot. Next, we compare the NN models with an empirical model in a climatological study. Finally, we analyze the prediction performance in a resolved temporal scale. The day-to-day electron density pattern prediction is shown for different models with a ranking study.

4.1 Determination of the optimal number of epochs through validation loss dips

In Table 3, the number of hidden layers and the number of neurons in each layer are shown. For the NAS models, these numbers were determined by the best validation loss from eight independent randomly initialized AutoKeras trainings. Since the early stop was used in NAS, a fine tune of learning rate was conducted using the training and validation loss curves where each tuning run consists of 8,000 epochs, after the network structures were determined. The training and validation loss curves for the best learning rate of each model (the last row of Table 3) are shown in Figure 3. As demonstrated, the validation loss curve floats slightly above the faster converging training loss and keeps decreasing until reaching the black dot. As the increase of the validation loss indicates the possibility of the model overfitting, we chose the epoch number as the dipping point.

***Table 3** The hyperparameters for four NN models, which are the optimal results of each category in architecture, learning rate, and validation loss dip epoch.*

	SLNN	DNN	SLNN-NAS	DNN-NAS
# of layers and neurons	[18]	[24, 22, 20]	[52]	[60, 32]
Learning rate	5e-04	9e-05	1.6e-04	7.7e-05
# of epochs	2195	4444	2116	6046

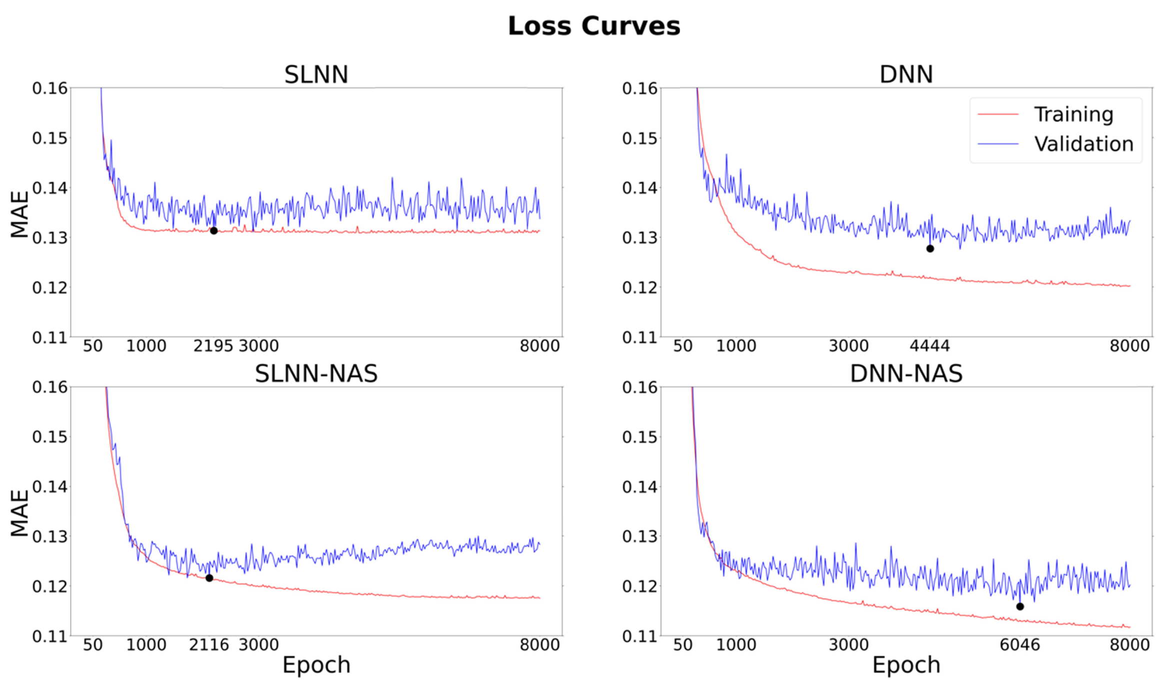


***Figure 3** The training (red) and validation (blue) loss curves of four NN models (the optimal number of epochs marked as the black dot). The two DNN models take more epochs to evolve the optimal results due to more complexity than SLNNs, while the NAS guided models lead to better model generality (lower possible validation loss).*

4.2 Overall prediction performance

Their quantitative metrics, MAE, RMSE, and RE, on the test data are shown in Table **4** below.

Table 4 *Prediction errors for four models in mean absolute error (MAE), root mean square error (RMSE), and relative error (RE) percentage.*

	SLNN	*DNN*	*SLNN-NAS*	*DNN-NAS*
MAE	0.1399	0.1312	0.1307	0.1250
RMSE	0.1908	0.1805	0.1821	0.1784
RE (%)	1.2667	1.1872	1.1844	1.1327

Two NAS models have lower prediction errors than their counterparts with fixed architectures. For example, NAS results in 6.6% reduction on MAE of *Ne* for SLNN and 4.7% reduction for DNN, respectively. DNN-NAS achieves the best prediction performance, i.e. lowest MAE, RMSE, and RE. Its improvement over SLNN is more than 10% on MAE and RE.

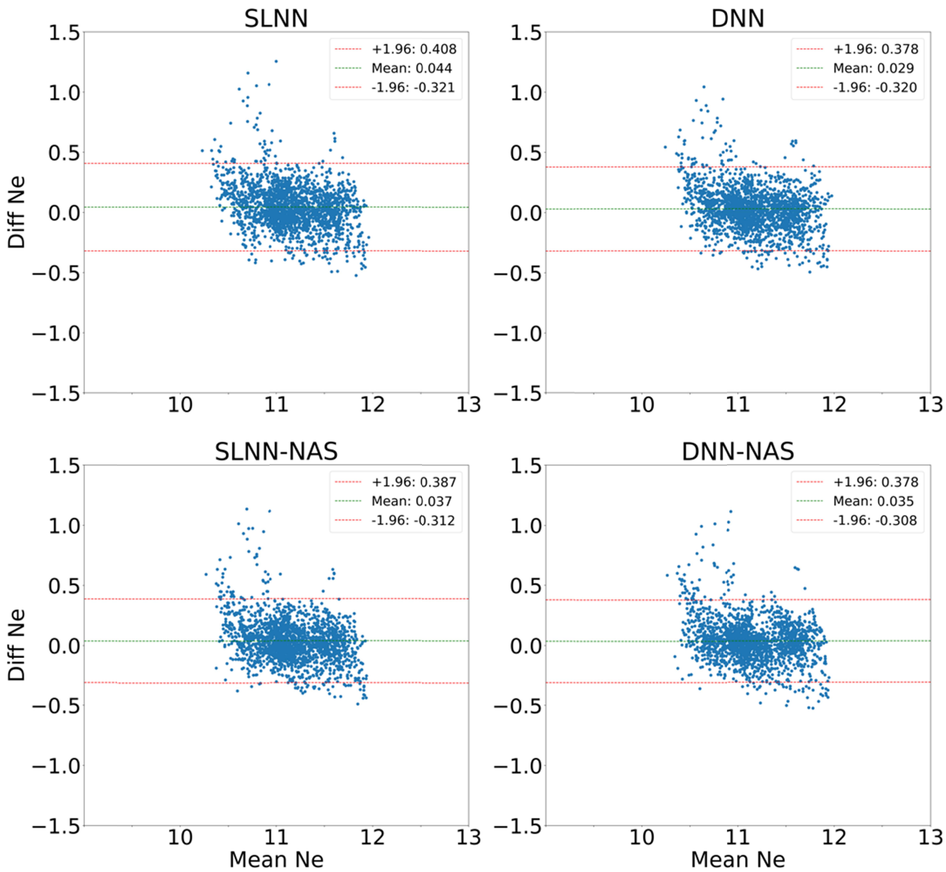


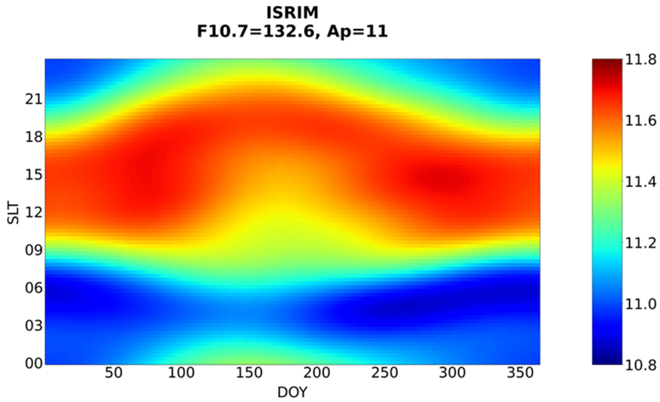
Figure 4 *BA-plots of the four optimal models (SLNN, DNN, SLNN-NAS, and DNN-NAS), in which the calculations are based on the test set. DNN tends to have the lowest averaged difference (green line in the upper right subplot) and the DNN-NAS owns the narrowest limits of agreements (distance between two red lines in the lower right subplot). The Y-axis is the Ne difference between the model prediction and the observation. The X-axis is the average of the model prediction and the observation.*

The Bland-Altman (BA) plots in Figure 4 show the agreement between each model prediction and the ground truth *Ne* from ISR observation. SLNN shows the least agreement with the largest bias and the widest 95% limits of agreement (± 1.96 SD). SLNN-NAS is better than SLNN, but still worse than DNN and DNN-NAS. DNN-NAS has a slightly larger bias but a narrower 95% limits of agreement than DNN. Again, DNN-NAS achieves the best agreement between the prediction and the ground truth since DNN-NAS adapts an optimal network structure and other hyperparameters, such as learning rate.

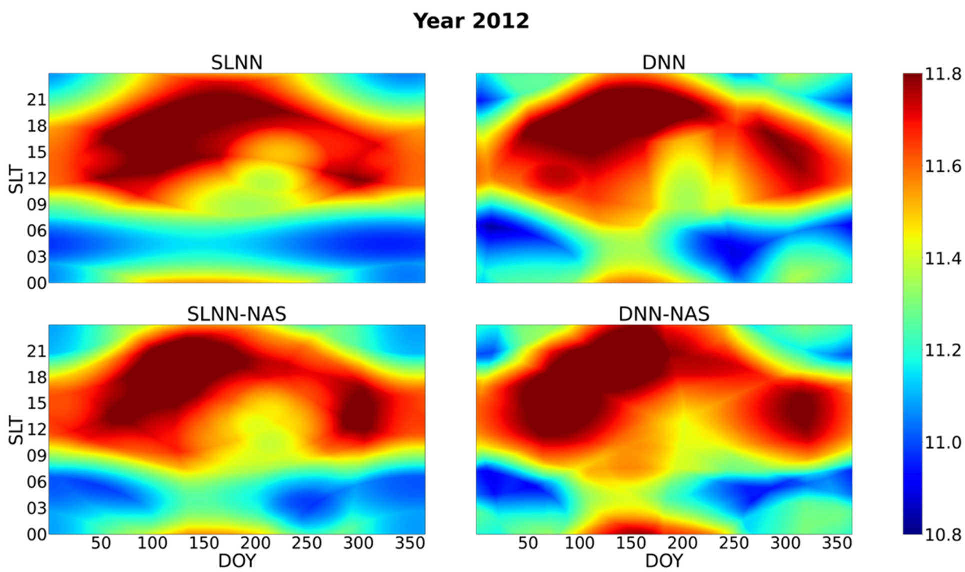
4.2 Climatological analysis

The climatological study can verify whether the NN models can keep track of *Ne* characteristics at a long temporal scale. For comparison, the ISRIM [*Holt et al.*, 2002; *Zhang and Holt*, 2007; *Zhang et al.*, 2005] was used, which is an open-source online tool for *Ne* climatological study under different conditions (altitude, geodetic latitude, F10.7, and Ap3). The annual *Ne* patterns from ISRIM (Figure 5 (a)) and four NN models (Figure 5 (b) and (c)) in 2012 are all plotted for 24 hours × 365 days (or 366 for the leap years). The temporal resolution of ISRIM is 18-minute which is practically the finest to achieve, while the temporal resolution of NN models are as fine as 4 minutes. Note that as ISRIM used the fixed altitude, F10.7, and Ap3, and the four NN models were run with the same fixed values to obtain Figure 5 (b). All NN models reproduce an asymmetric semi-annual pattern of *Ne* as shown in ISRIM, which resembles as a saddle-like structure with *Ne* concentration peaks in Spring and Fall. The two SLNN models show more choppy edges on the crests, which could imply the incapability of the simple architecture to fully catch the data characteristics. DNN-NAS seems to have two more appealing crests, while the other three NN models suffer a star like artifact at the center. Furthermore, the NN models provide a detailed prediction (Figure 5 (c)) to fill the limited observation (Figure 1), using real-time F10.7 and Ap3. The 14 isolated thread-like enhancements in Figure 5 **Error! Reference source not found.**(c) could be the indication of 27-day mid-latitude topside ionospheric electron variation [*Rich et al.*, 2003].

(a) ISRIM climatological pattern of medium solar activity.



(b) semi-annual patterns of climatological study.



(c) semi-annual patterns based on external geophysical indices.

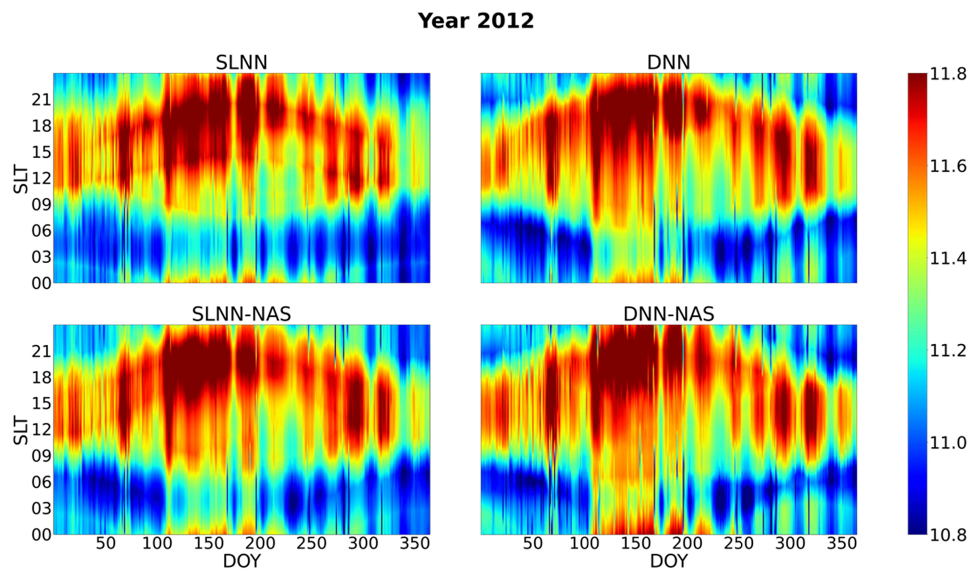
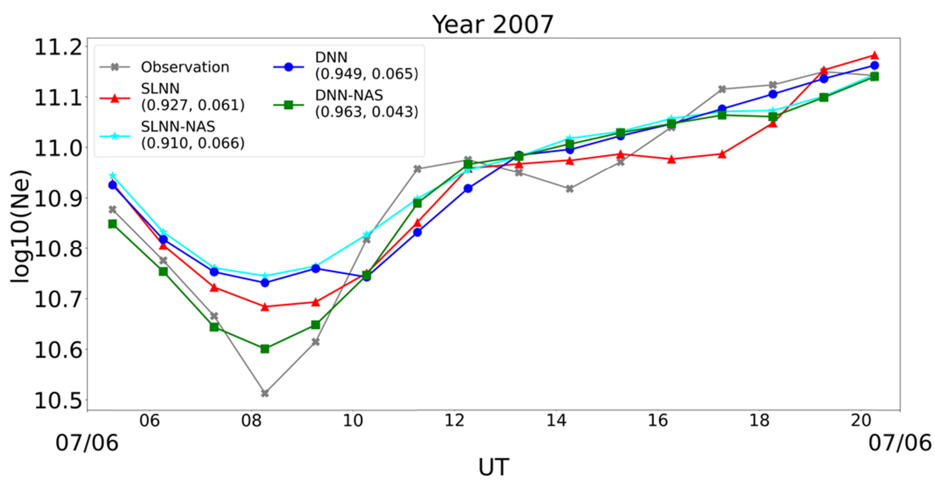


Figure 5 *Annual electron density patterns of year 2012 from different sources: (a) ISR empirical model (ISRIM), (b) four model predictions based on the fixed F10.7 and Ap3, (c) four model predictions based on the real-time F10.7 and Ap3. Based on the nature of neural network models, the input can be arbitrary values. We set the evenly distributed temporal information to get the time related drivers (year, DOY, and SLT), while comparison between (a) and (b) serves as the comparison on the climatological study, while (c) demonstrates a more realistic case of Ne annual pattern with real-time F10.7 and Ap3 inputs.*

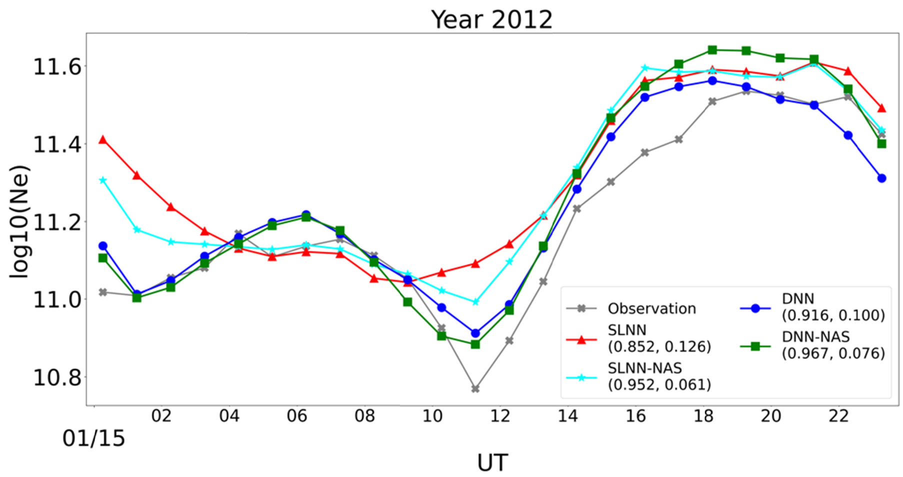
4.3 Daily *Ne* pattern prediction

To evaluate the model performance, the daily *Ne* patterns were compared to illustrate how well the models predict in a resolved temporal scale varying from annual to daily. All the drivers (year, cyclic DOY and cyclic SLT, F10.7, and Ap3) served to get the model output. The two geophysical indices were obtained from OpenMadrigal database of MIT Haystack Observatory if not available in ISR. Since the days with full hourly *Ne* coverage are limited in the ISR data, we have identified a total of 128 days in the test data with a decent full-day hourly coverage. Three examples of hourly changes of *Ne* in a day (07/06/2007, 01/15/2012, and 08/01/2012) are shown in Figure 6 with observations and different model outputs. Note that 07/06/2007 and 08/01/2012 do not have a full 24-hour coverage. To quantify the agreement between the prediction and the observation, Pearson correlation coefficients (CCs) and MAEs are calculated and shown in Figure 6. The higher CC values indicate the better trend match (with the removal of the mean and normalization) and the lower MAEs indicate less discrepancies between prediction and observation. In general, all NN models follow the observation patterns (gray cross) well and DNN-NAS achieves the largest CC (and the smallest MAE except for 01/15/2012). For 01/15/2012 in Figure 6 (b), the dip is later than the other two cases since the sun rises later in winter than in summer. Figure 6 shows that DNN-NAS predicts the observations better than the other three models, which are the dominant cases in all 128 days with a good daily coverage.

(a) 2007-07-06



(b) 2012-01-15



(c) 2012-08-01

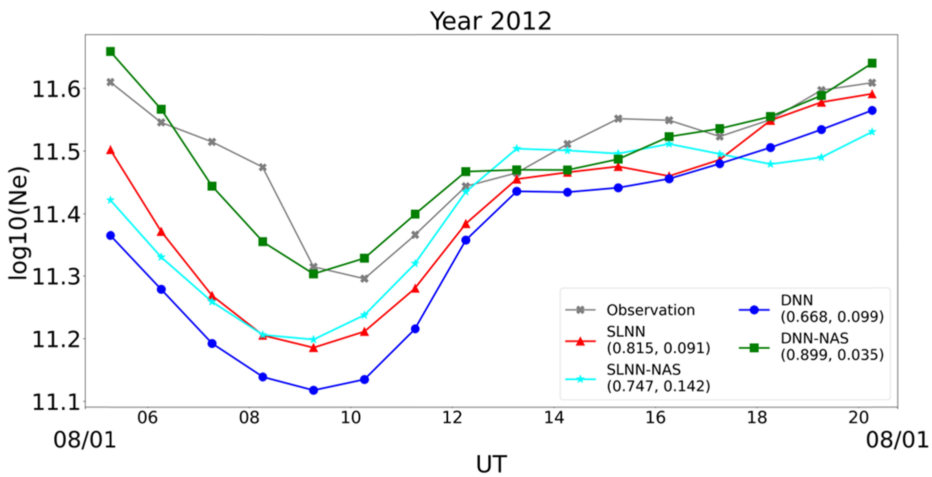


Figure 6 *Daily Ne pattern prediction on three different days: (a) 2007-07-06, (b) 2012-01-15, and (c) 2012-08-01. Gray cross: the ISR observation; red triangle: SLNN; cyan star: SLNN-NAS; blue circle: DNN; green square: DNN-NAS. The two parameters (Pearson correlation coefficients and MAE) help evaluate how well model outputs predict the observed diurnal Ne pattern. Generally, all model outputs follow the observed diurnal pattern well, while DNN-NAS predicts the best.*

We calculated CC and MAE for all 128 daily patterns from the test set and ranked four models. The number of ranks for four models are shown in Table 5. Specifically, 1-4 ranks are corresponding to the decreasing CC or increasing MAE. For example, rank 1 represents the largest CC or the least MAE, which corresponds the best prediction of daily pattern. And rank 4 represents the smallest CC or the largest MAE, which corresponds the worst prediction of daily pattern. As can be seen, DNN-NAS has a dominatingly good prediction performance with 61 (48%) for CC (rank #1) and 54 (42%) for MAE (rank #1).

Table 5 *The ranks for daily pattern predictions. Among the 128 days in the test set, the Pearson correlation coefficients (CCs) and mean absolute errors (MAEs) are calculated and sorted from best (highest CC or lowest MAE). The DNN-NAS shows the greatest number of rank 1 cases.*

		SLNN	**DNN**	**SLNN-NAS**	**DNN-NAS**
CC	Rank 1	25	16	26	61 (48%)
	Rank 2	26	35	41	26
	Rank 3	32	48	31	17
	Rank 4	45	29	30	24
MAE	Rank 1	17	30	27	54 (42%)
	Rank 2	34	32	33	29
	Rank 3	29	32	44	23
	Rank 4	48	34	24	22

To assess the performance of NN models during several continuous days, the duration with decent observation coverage is selected for a further comparison. Two indices as drivers (F10.7 and Ap3) are shown in the upper panel of Figure 7. A cubic interpolation is applied to the Ap3 index for the reference purpose. In Figure 7, both model predictions and observations show a strong correlation to Ap3. When the Ap3 index increases from quiet time to moderate active value, the increase on Ap3 tends to cause a decrease in *Ne* at the three post-midnights from September 3rd to September 5th, which indicates a negative ionospheric storm phase. All NN models seem to track these changes well, while DNN-NAS seems to track the observation best (with the highest CC and the second lowest MAE).

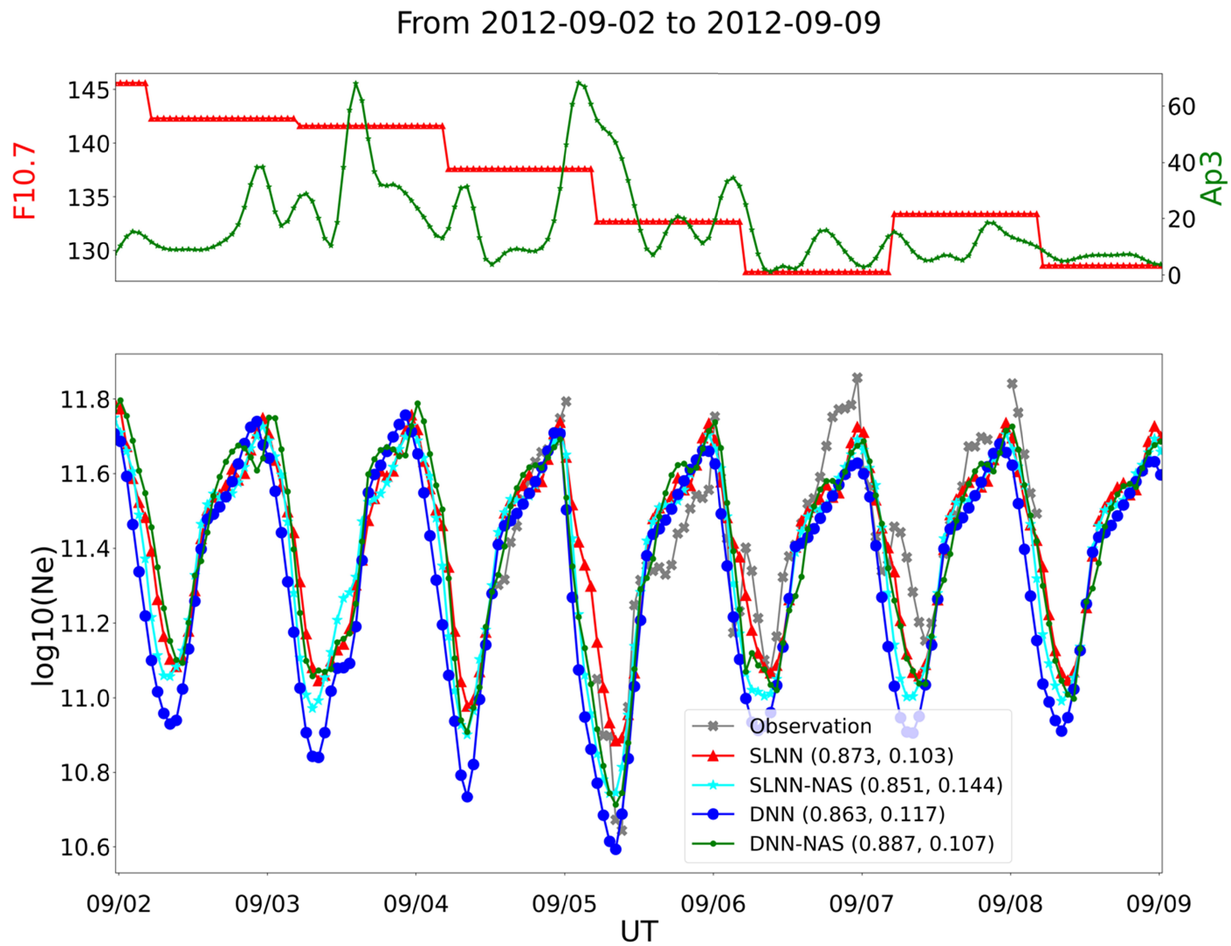


Figure 7 *Ne patterns during 2012-09-02 to 2012-09-09. The two geophysical drivers are drawn in the upper panel. Four model outputs are of different markers followed with CCs and MAEs (based on observational values) in parentheses. Clearly, we see the Ap3 serves as the major driver effect to the model outputs as the predictions dip down when Ap3 reaches its peak at early time of September 5th.*

5 Discussion

In this study, we have shown that NAS helps find an optimal neural network setting to reduce the *Ne* prediction errors for both SLNN and DNN. Furthermore, NAS could make the process more efficient with little manual interventions. Generally, we started with a large and sparse search poll of assigned hyperparameters. Based on the behavior of loss curves, the search poll was refined to reach the optimal neural architect. The multiple GPU cores facilitated this hierarchic search. The manual determination of the optimal network structure is a daunting work. For example, with a fixed number of three hidden layers, thousands of full trainings were performed to obtain the number of neurons in each layer [*Li et al.*, 2021]. Even the simple selection of the optimal learning rate could involve a substantial amount of manual work as we did for the two manual models. The NAS provides an efficient way to identify the optimal hyperparameters for NN models. For the current simple application of NAS for *Ne* prediction at the fixed geophysical location and altitude, the search process is fast (about 33 minutes on NVIDIA A6000, 22 minutes for NAS search and 11 minutes for additional epochs). However, the converging status of training and validation curves is absent in the early-stopping search. Considerate amount of manual work is still required to run additional epochs based on the NAS guided architectures and analyze the loss curves. Thus, we would assume more advanced NAS application could further reduce the tedious work spent in optimizing the neural networks.

Overfitting remains a general concern with machine learning models. As shown in Figure 3, the training loss could be continuously reduced. As a matter of fact, when we used a complex NN model, the fitting error can approach a very low value at the cost of reducing model generalization to an acceptable level with high prediction errors. Thus, the validation dip in Figure 3 alleviates this issue. Furthermore, NAS uses an early-stopping criterion for an efficient search. For highly nonlinear problems, NAS could trap in a local minimum. We used multiple random initializations for NAS to avoid this problem. DNN-NAS stands out in the overall quantitative measurements, climatological study, and prediction rankings of daily patterns.

All NN models predict *Ne* well during the moderate event in the daytime section (Figure 7). This is consistent with previous studies of *Ne* prediction using NN models and due to a couple of reasons. First, the training data are confined to the condition (Ap3 ≤ 80 in Table 2), which causes the NN models to be prone to these cases. Secondly, the physical drivers are not fine enough in time, e.g. F10.7 is a daily average and Ap3 is 3-hour average. We conducted an additional training of DNN-NAS without the restriction on Ap3 (i.e, Ap3 could be larger than 80 which covers intense storm periods), namely DNN-NAS*. The comparison between DNN-NAS and DNN-NAS* is shown in Figure 8. The shade region is approximately from 05UT to 15UT on March 9th, 2012. Though DNN-NAS has overall better CC and MAE, DNN-NAS* showed a much larger CC and lower MAE than DNN-NAS in the shade region. However, both models struggle to track the *Ne* dip around 08UT on March 10th. As the ISR data with Ap3 ≥ 80 are only account for less than 2% of the total data, it is not a surprise that DNN-NAS* only improved over DNN-NAS in certain regions and suffered performance loss in other regions. In future work, either a separate model for major geomagnetic events or a general model with different weights on these events should be built with more event data to address this challenging problem.

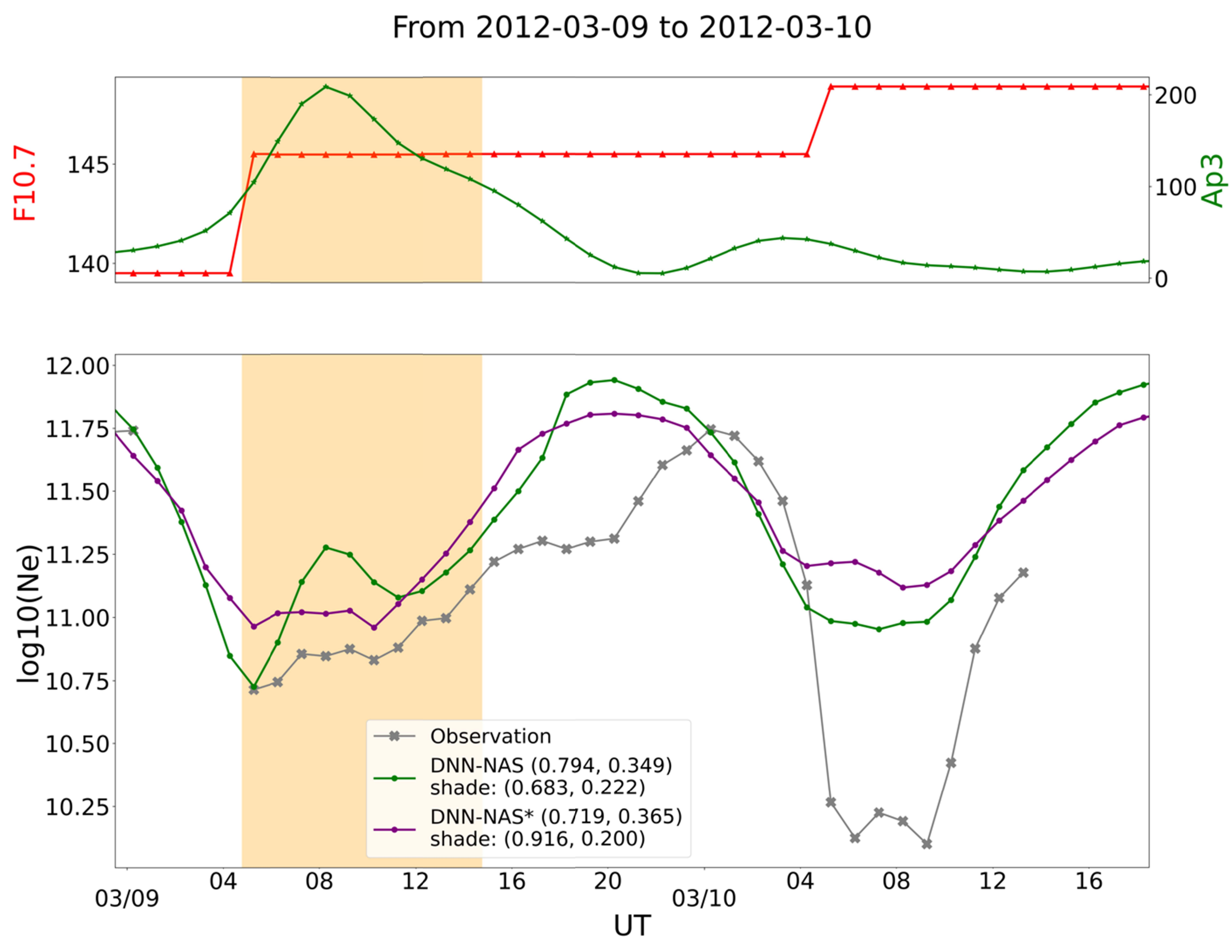


Figure 8 *DNN-NAS trained with Ap3≤80 and DNN-NAS* trained without the restriction on Ap3., the DNN-NAS models trained with and without filter on Ap3 have the prediction results in green and purple color. The CC and MAE calculated on the observational data are in the parentheses (the whole curve after the model name and the shade region after "shade").*

One possible reason for the limited performance improvement of NAS models over fixed NN models may be the lack of sufficient training data. To address this issue, we applied cubic B-spline to the vertical profiles of *Ne* with 15-minute cadence. After removing the abnormal data points, a total of nearly 43,000 data points around 350 km were used for training/validation/test (where the test set was changed to 2007 and 2016 in order to balance the amount of validation and test data), about 4 times of data points with 1-hr cadence (11,483). In the following sections, we call the data with 1-hr cadence as the 1-hr dataset (2007 and 2012 as test data) and that with

15-minute cadence as the 15-min dataset (2007 and 2016 as test data). However, the models trained on the 15-min dataset led to similar findings as those using the 1-hr dataset, i.e. NAS led to only marginal improvement of *Ne* prediction. Therefore, the lack of sufficient training data may not be the primary reason for the limited improvement of NAS models.

We further conducted a complexity analysis of NN by increasing the number of network weights of a SLNN (denoted as "Complexity") on the 15-min dataset. The validation loss is plotted with the change of the complexity of SLNN in Figure 9. As can be seen, the loss function drops quickly at the beginning and converges to a steady level slightly below 0.125 after the complexity reaches 128 network parameters, i.e. 16 neurons in the hidden layer. Therefore, for the ISR *Ne* data, fully connected NN seems to reach its performance limit at a simple structure. This explains why NAS could only achieve limited improvement over fixed NN models, which are already complex enough to model the data in hand.

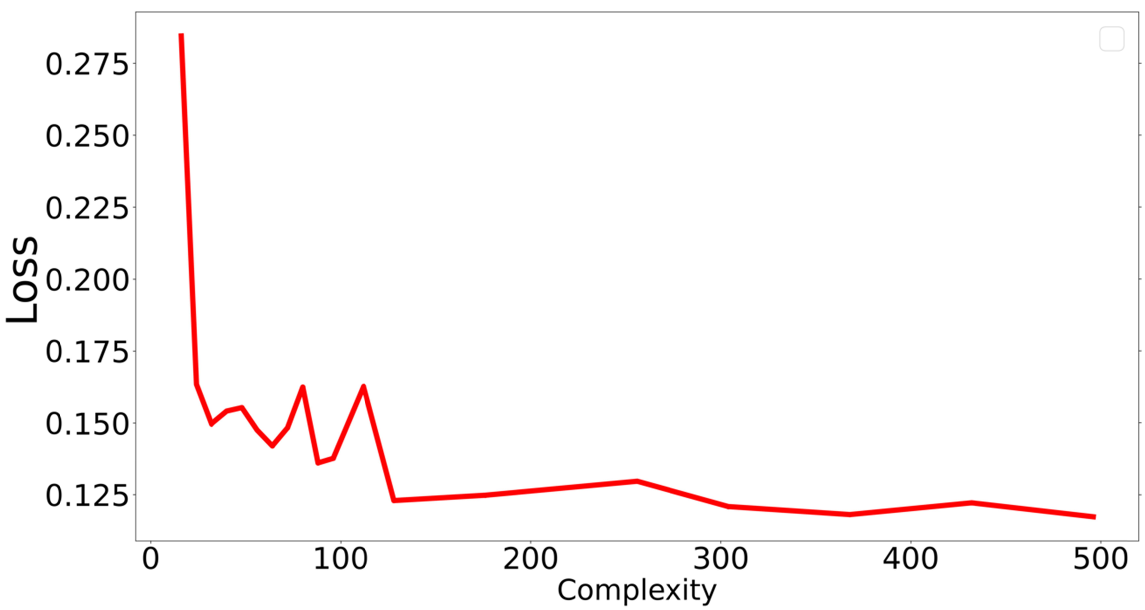
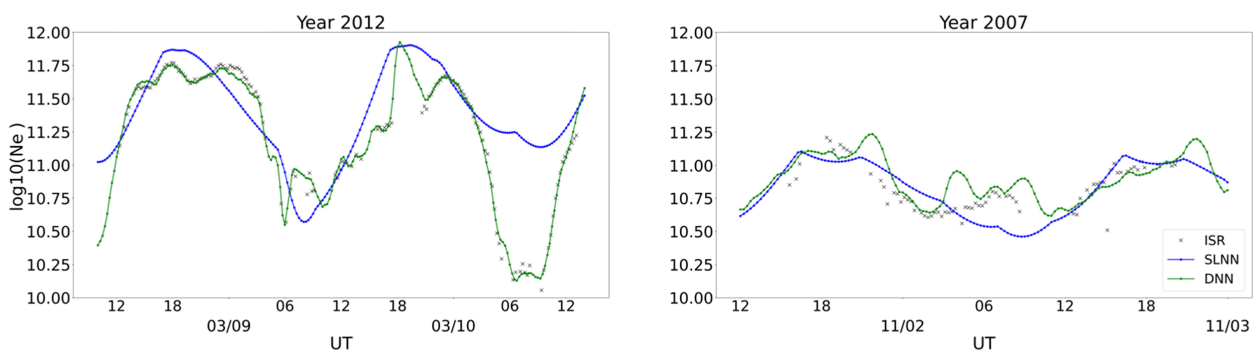


Figure 9 *Prediction performance changes along with the model complexity. The complexity is defined as the total number of trainable weights of the NN model. The mean absolute error of the validation set serves as the loss function, where the less loss indicates the better performance.*

It is also worth noting that DNN-NAS could achieve much better fitting, but at the cost of losing the generality. For an overtrained DNN-NAS model ([512, 512, 512, 512, 32]), the training MAE (after 7,900 epochs of training) is as low as 0.0529, compared to 0.1261 of SLNN for the 15-min dataset. However, the test MAE dropped to 0.1587, compared to 0.1285 of SLNN, which indicates the loss of generality of DNN-NAS. In Figure 10 (a) for the training data (2012 for the 15-min dataset), the overtrained DNN-NAS can fit the complicated structures on March 9th and dip on March 10th of the observations, while SLNN fails to catch these structures. However, in Figure 10 (b) for the test data (2007 for the 15-mind dataset), DNN-NAS shows some abnormal oscillations as the signs of overfitting.



(a) (b)

Figure 10 *Overfitting of DNN (architecture: [512, 512, 512, 512, 32], green) (a) fitting and (b) prediction. SLNN (18 hidden neuron, blue) is served as a benchmark. DNN can fit the ISR data more closely than SLNN as shown in (a). However, DNN leads to an unrealistic wavy pattern for prediction as shown in (b).*

Finally, the current study confines to *Ne* prediction at a fixed latitude and altitude in order to investigate the effectiveness of different NN models. 3D NN models have been proposed using the ionospheric radio occultation measurements in previous studies [*Gowtam et al.*, 2019; *Habarulema et al.*, 2021]. The static nature of fully connected NN is also accountable for the limited prediction performance of this study (in line with previous studies) as electron density change is a dynamic process, influenced by different geomagnetic parameters or other factors at different space and time scales. For example, the increase of Ap3 affects neutral density, which can cause the electron density change over the next few hours rather than the instant change. Though the geophysical indices serve as the drivers in many developed models [*XN Chu et al.*, 2017; *X Chu et al.*, 2017; *Habarulema et al.*, 2021; *Li et al.*, 2021], the atmospheric neutral components at Millstone Hills, which have shown strong correlations with electron density, may not be accurately described by the current input parameters of the NN models (F10.7 and Ap3). Technically, the more advanced generative models with the time histories of the input parameters may lead to much more improved prediction than the fully connected NN models without memory mechanism. Besides, this study examined the feasibility of applying NAS in identifying an optimal network structure of future works on either building electron density vertical profile based on ISR or other electron density models. Combined with aforementioned technical advancement, electron density prediction offered by deep learning could be significantly improved. And new drivers may be needed to accommodate the resolved temporal resolution, such as adding the 81-day average F10.7 (F10.7p) for the historical information or the geomagnetic AE index, and the physical processes, such as neutral composition, in our future work. Last but not least, information theory can help identify and select the drivers and their time histories that are relevant for predicting the output parameter, e.g., solar wind parameters [*Simon Wing et al.*, 2016; *Simon Wing et al.*, 2022a; *Simon Wing et al.*, 2022b].

6 Conclusion

We demonstrate that neural architecture search (NAS) that can identify the optimal network structure automatically for *Ne* prediction at a fixed height using 16-year ISR observations at

Millstone Hill. In addition to modeling efficiency, NAS derived DNN models also lead to better prediction performance than manually tuned SLNN (more than 10% improvement on MAE and RE) and rank the highest for daily *Ne* pattern prediction based on CC and MAE. The climatological *Ne* patterns from different NN models reveal the two crests in Spring and Fall seasons in general. We also investigated the reason for limited improvement of NAS due to the network complexity and the lack of memory mechanism of the fully connected NN. In future, the more advanced generative models with a memory mechanism and better resolved and understood physical drivers of these models will be pursued for a much-improved 3D *Ne* prediction.

Acknowledgment

This research conducted at the University of Texas at Arlington was supported by AFOSR through award FA9550-23-1-0634 and NASA under the grant numbers of 80GSFC22CA011 and 80NSSC20K1786. We thank the support from the ISSI for the international team on "Multi-Scale Magnetosphere-Ionosphere-Thermosphere Interaction".

Data Availability Statement

The ISR-NAS models and data to plot the figures in this study are available at https://doi.org/10.5281/zenodo.8350762.

References

Athieno, R., P. T. Jayachandran, and D. R. Themens (2017), A neural network-based foF2 model for a single station in the polar cap, *Radio Science*, *52*(6), 784-796.

Baker, B., O. Gupta, N. Naik, and R. Raskar (2016), Designing neural network architectures using reinforcement learning, *arXiv preprint arXiv:1611.02167*.

Bilitza, D. (2001), International reference ionosphere 2000, *Radio science*, *36*(2), 261-275.

Bilitza, D., D. Altadill, V. Truhlik, V. Shubin, I. Galkin, B. Reinisch, and X. Huang (2017), International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions, *Space weather*, *15*(2), 418-429.

Bortnik, J., W. Li, R. Thorne, and V. Angelopoulos (2016), A unified approach to inner magnetospheric state prediction, *Journal of Geophysical Research: Space Physics*, *121*(3), 2423-2430.

Cai, H., L. Zhu, and S. Han (2018a), Proxylessnas: Direct neural architecture search on target task and hardware, *arXiv preprint arXiv:1812.00332*.

Cai, H., T. Chen, W. Zhang, Y. Yu, and J. Wang (2018b), Efficient architecture search by network transformation, paper presented at Proceedings of the AAAI Conference on Artificial Intelligence.

Cai, Y., X. Yue, W. Wang, S. Zhang, L. Liu, H. Liu, and W. Wan (2019), Long-term trend of topside ionospheric electron density derived from DMSP data during 1995–2017, *Journal of Geophysical Research: Space Physics*, *124*(12), 10708-10727.

Chu, X., J. Bortnik, W. Li, Q. Ma, V. Angelopoulos, and R. Thorne (2017), Erosion and refilling of the plasmasphere during a geomagnetic storm modeled by a neural network, *Journal of Geophysical Research: Space Physics*, *122*(7), 7118-7129.

Chu, X., J. Bortnik, W. Li, Q. Ma, R. Denton, C. Yue, V. Angelopoulos, R. Thorne, F. Darrouzet, and P. Ozhogin (2017), A neural network model of three-dimensional dynamic electron density in the inner magnetosphere, *Journal of Geophysical Research: Space Physics*, *122*(9), 9183-9197.

Desell, T. (2017), Large scale evolution of convolutional neural networks using volunteer computing, paper presented at Proceedings of the Genetic and Evolutionary Computation Conference Companion.

Elsken, T., J.-H. Metzen, and F. Hutter (2017), Simple and efficient architecture search for convolutional neural networks, *arXiv preprint arXiv:1711.04528*.

Elsken, T., J. H. Metzen, and F. Hutter (2019), Neural architecture search: A survey, *The Journal of Machine Learning Research*, *20*(1), 1997-2017.

Gowtam, V. S., S. Tulasi Ram, B. Reinisch, and A. Prajapati (2019), A new artificial neural network-based global three-dimensional ionospheric model (ANNIM-3D) using long-term ionospheric observations: Preliminary results, *Journal of Geophysical Research: Space Physics*, *124*(6), 4639-4657.

Guo, Z., X. Zhang, H. Mu, W. Heng, Z. Liu, Y. Wei, and J. Sun (2020), Single path one-shot neural architecture search with uniform sampling, paper presented at European conference on computer vision, Springer.

Habarulema, J. B., D. Okoh, D. Burešová, B. Rabiu, M. Tshisaphungo, M. Kosch, I. Häggström, P. J. Erickson, and M. A. Milla (2021), A global 3-D electron density reconstruction model based on radio occultation data and neural networks, *Journal of Atmospheric and Solar-Terrestrial Physics*, *221*, 105702.

Holt, J. M., S.-R. Zhang, and M. J. Buonsanto (2002), Regional and local ionospheric models based on Millstone Hill incoherent scatter radar data, *Geophysical Research Letters*, *29*(8), 48-41-48-43.

Hutter, F., L. Kotthoff, and J. Vanschoren (2019), *Automated machine learning: methods, systems, challenges*, Springer Nature.

Jin, H., Q. Song, and X. Hu (2019), Auto-keras: An efficient neural architecture search system, paper presented at Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining.

Kingma, D. P., and J. Ba (2014), Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.

Lei, J., S. Syndergaard, A. G. Burns, S. C. Solomon, W. Wang, Z. Zeng, R. G. Roble, Q. Wu, Y. H. Kuo, and J. M. Holt (2007), Comparison of COSMIC ionospheric measurements with ground-based observations and model predictions: Preliminary results, *Journal of Geophysical Research: Space Physics*, *112*(A7).

Li, W., D. Zhao, C. He, Y. Shen, A. Hu, and K. Zhang (2021), Application of a Multi-Layer Artificial Neural Network in a 3-D Global Electron Density Model Using the Long-Term Observations of COSMIC, Fengyun-3C, and Digisonde, *Space Weather*, *19*(3), e2020SW002605.

Liu, L., W. Wan, B. Ning, O. Pirog, and V. Kurkin (2006), Solar activity variations of the ionospheric peak electron density, *Journal of Geophysical Research: Space Physics*, *111*(A8).

Luo, R., F. Tian, T. Qin, E. Chen, and T.-Y. Liu (2018), Neural architecture optimization, *Advances in neural information processing systems*, *31*.

Real, E., S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin (2017), Large-scale evolution of image classifiers, paper presented at International Conference on Machine Learning, PMLR.

Rich, F. J., P. J. Sultan, and W. J. Burke (2003), The 27-day variations of plasma densities and temperatures in the topside ionosphere, *Journal of Geophysical Research: Space Physics*, *108*(A7).

Richards, P., D. Torr, B. Reinisch, R. Gamache, and P. Wilkinson (1994), F 2 peak electron density at Millstone Hill and Hobart: Comparison of theory and measurement at solar maximum, *Journal of Geophysical Research: Space Physics*, *99*(A8), 15005-15016.

Sai Gowtam, V., and S. Tulasi Ram (2017), An Artificial Neural Network-Based Ionospheric Model to Predict NmF2 and hmF2 Using Long-Term Data Set of FORMOSAT-3/COSMIC Radio Occultation Observations: Preliminary Results, *Journal of Geophysical Research: Space Physics*, *122*(11), 11,743-711,755.

Suganuma, M., S. Shirakawa, and T. Nagao (2017), A genetic programming approach to designing convolutional neural network architectures, paper presented at Proceedings of the genetic and evolutionary computation conference.

Tulasi Ram, S., V. Sai Gowtam, A. Mitra, and B. Reinisch (2018), The improved two-dimensional artificial neural network-based ionospheric model (ANNIM), *Journal of Geophysical Research: Space Physics*, *123*(7), 5807-5820.

Wing, S., J. R. Johnson, E. Camporeale, and G. D. Reeves (2016), Information theoretical approach to discovering solar wind drivers of the outer radiation belt, *Journal of Geophysical Research: Space Physics*, *121*(10), 9378-9399.

Wing, S., J. R. Johnson, D. L. Turner, A. Y. Ukhorskiy, and A. J. Boyd (2022a), Untangling the solar wind and magnetospheric drivers of the radiation belt electrons, *Journal of Geophysical Research: Space Physics*, *127*(4), e2021JA030246.

Wing, S., D. L. Turner, A. Y. Ukhorskiy, J. R. Johnson, T. Sotirelis, R. Nikoukar, and G. Romeo (2022b), Modeling radiation belt electrons with information theory informed neural networks, *Space Weather*, *20*(8), e2022SW003090.

Wing, S., J. Johnson, J. Jen, C. I. Meng, D. Sibeck, K. Bechtold, J. Freeman, K. Costello, M. Balikhin, and K. Takahashi (2005), Kp forecast models, *Journal of Geophysical Research: Space Physics*, *110*(A4).

Yang, D., and H. Fang (2023), A Low-Latitude Three-Dimensional Ionospheric Electron Density Model Based on Radio Occultation Data Using Artificial Neural Networks With Prior Knowledge, *Space Weather*, *21*(1), e2022SW003299.

Yue, X., L. Hu, Y. Wei, W. Wan, and B. Ning (2018), Ionospheric Trend Over Wuhan During 1947–2017: Comparison Between Simulation and Observation, *Journal of Geophysical Research: Space Physics*, *123*(2), 1396-1409.

Zhang, S. R., and J. M. Holt (2007), Ionospheric climatology and variability from long-term and multiple incoherent scatter radar observations: Climatology in eastern American sector, *Journal of Geophysical Research: Space Physics*, *112*(A6).

Zhang, S. R., J. M. Holt, A. P. Van Eyken, M. McCready, C. Amory-Mazaudier, S. Fukao, and M. Sulzer (2005), Ionospheric local model and climatology from long-term databases of multiple incoherent scatter radars, *Geophysical Research Letters*, *32*(20).

Zoph, B., and Q. V. Le (2016), Neural architecture search with reinforcement learning, *arXiv preprint arXiv:1611.01578*.