

# Space Weather®

## RESEARCH ARTICLE

10.1029/2023SW003505

### Special Section:

Machine Learning in  
Heliophysics

### Key Points:

- Clusters of radar echo bins associated with Equatorial Spread-*F* can be identified using the DBSCAN algorithm from coherent scatter data
- A random forest model showed that the day of year and vertical plasma drifts previous to sunset are good predictors of high SNR clusters
- A convolutional neural network was able to predict ionospheric irregularities with an accuracy of more than 70% despite scarce data

### Correspondence to:

D. Garcia,  
dg565@cornell.edu

### Citation:

Garcia, D., Rojas, E. L., & Hysell, D. L. (2023). Predicting equatorial ionospheric convective instability using machine learning. *Space Weather*, 21, e2023SW003505. <https://doi.org/10.1029/2023SW003505>

Received 28 MAR 2023

Accepted 31 OCT 2023

### Author Contributions:

**Conceptualization:** D. Garcia, E. L. Rojas, D. L. Hysell  
**Data curation:** D. Garcia, D. L. Hysell  
**Formal analysis:** D. Garcia, E. L. Rojas, D. L. Hysell  
**Investigation:** D. Garcia, D. L. Hysell  
**Methodology:** D. Garcia, E. L. Rojas, D. L. Hysell  
**Software:** D. Garcia, D. L. Hysell  
**Supervision:** E. L. Rojas, D. L. Hysell  
**Validation:** D. Garcia  
**Visualization:** D. Garcia, E. L. Rojas, D. L. Hysell  
**Writing – original draft:** D. Garcia, E. L. Rojas, D. L. Hysell  
**Writing – review & editing:** D. Garcia, E. L. Rojas, D. L. Hysell

© 2023. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Predicting Equatorial Ionospheric Convective Instability Using Machine Learning

D. Garcia<sup>1</sup> , E. L. Rojas<sup>2</sup> , and D. L. Hysell<sup>2</sup> 

<sup>1</sup>Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA, <sup>2</sup>Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA



**Abstract** The numerical forecast methods used to predict ionospheric convective plasma instabilities associated with Equatorial Spread-*F* (ESF) have limited accuracy and are often computationally expensive. We test whether it is possible to bypass first-principle numeric simulations and forecast irregularities using machine learning models. The data are obtained from the incoherent scatter radar at the Jicamarca Radio Observatory located in Lima, Peru. Our models map vertical plasma drifts, time, and solar activity to the occurrence and location of clusters of echoes telltale of ionospheric irregularities. Our results show that these models are capable of identifying the predictive power of the tested inputs, obtaining accuracies around 75%.

**Plain Language Summary** Equatorial Spread-*F* (ESF) is a phenomenon that happens in the ionosphere and it can create disturbances that affect communication and navigation systems, such as GPS or satellite phones. Scientists have been trying to predict when and where ESF will happen using numerical forecast methods, but these methods are not very accurate and can be expensive. In this study, we tested a new way to predict ESF using machine learning models. We used data from a radar in Lima, Peru, and looked at information such as how the plasma in the ionosphere moves vertically and how active the sun was. The models were able to predict where and when ESF would occur with about 75% accuracy, which is a promising development.

### 1. Introduction

A signature feature of space weather in the postsunset equatorial ionosphere is the frequent occurrence of deep, broadband plasma density irregularities driven by plasma convective instability. The irregularities can occupy scale sizes from tens of cm through hundreds of km, span altitudes from the valley region to the topside, and persist for hours. They alter radio wave reflection, refraction, diffraction, and scattering in the ionosphere and, consequently, can disrupt a range of radio communication, navigation, and imaging systems. Plasma convective instability is responsible for the so-called “equatorial spread *F*” or “ESF” which has become an umbrella term for the phenomenon. Extensive reviews of the subject have been presented in the literature (see e.g., Kelley et al., 2011; Kil, 2015; Woodman, 2009).

The global climatology of ionospheric irregularities associated with ESF is now well established (see e.g., Chapagain et al., 2009; Gentile et al., 2006). Although ESF was discovered more than 80 years ago (Booker & Wells, 1938), forecasts models more reliable than climatologically based models remain elusive. An important step toward forecasting was taken more than 50 years ago when Farley et al. (1970) pointed out that there seemed to be a threshold *F*-region layer height required for instability. This highlighted the critical role of the background zonal electric field (vertical plasma drift), which influences layer height around sunset and can both establish the conditions for and subsequently drive instability (e.g., Fejer & Kelley, 1980). It is well known that the background zonal electric field reverses direction around sunset and exhibits a characteristic “prereversal enhancement,” or PRE, just before doing so. Understanding the PRE, modeling it, and estimating the impact of PRE on ionospheric convective instability are key areas of contemporary research into the forecast problem (Chen et al., 2017; Eccles et al., 2015; Farley et al., 1986; Fejer et al., 1979; Kil et al., 2009; Richmond et al., 2015).

Electric fields and plasma drifts can be measured accurately near the geomagnetic equator using the incoherent scatter radar (ISR) technique (Woodman, 1970). A substantial database has been accumulated with the Jicamarca Radio Observatory’s main radar, located outside Lima, Peru, near the magnetic equator. The database includes vertical plasma drift measurements and plasma density irregularities associated with ESF. The database, therefore, includes both causes and effects. Efforts have been underway for over a decade to use the data to develop

and test direct numerical simulations of plasma convective instability using these data (see Hysell et al. (2022) and references therein). While the results are promising and suggest that the time series of the vertical plasma drifts around sunset seems to be the main ingredient required for local irregularity forecasting, setting up and running each individual simulation is very time-consuming. Only a small fraction of the database will ever be investigated this way.

In this paper, we attempt to bypass first-principles numerical simulations and test whether it is possible to forecast ionospheric irregularities using vertical plasma drift measurements and machine learning. The project has three components. The first is to develop a way to quantify the degree of instability observed at Jicamarca on a given night from the ISR data. The second is to quantify the causal relationship between the drift data and subsequent instability. The third is to make and assess forecasts. Below, we describe the ISR data set from Jicamarca and then address each component. We conclude by summarizing with an assessment of the forecast potential of our methods.

## 2. Measurements and Data Sets Overview

Here, we describe the measurements and data used to classify and forecast ionospheric plasma irregularities using machine learning. We also describe how the features and labels were selected to feed the machine-learning model. Furthermore, we outline how the “density-based spatial clustering of applications with noise” or DBSCAN algorithm was used to label the observed instabilities automatically. Finally, we describe the metrics used to filter high-quality data samples.

### 2.1. Input Data: Vertical Drifts, Day of Year, and Solar Activity

The data set used in this study consists of equatorial vertical and east-west plasma drift measurements made by the ISR of the Jicamarca Radio Observatory ( $11.95^{\circ}\text{S}$ ,  $76.87^{\circ}\text{W}$ ,  $1^{\circ}$  dip) between the years 2010 and 2021. The drifts are measured using radar beams directed nearly perpendicular to the geomagnetic field. This facilitates accurate drift measurements and permits the detection of coherent scatter from field-aligned density irregularities when present (Kudeki et al., 1999). The Jicamarca ISR provides 10–45 days of data acquired in this drifts mode each year.

The drift mode has a nominal altitude resolution of 15 km and produces measurements from altitudes between 200 and 1,000 km with a time resolution of 5 min. Consequently, the natural way to visualize vertical backscatter radar data is as a 2-dimensional array with parameters represented through a color map. The parameters of interest are the ion drift velocities, measured in meters per second, and the signal-to-noise power ratio (SNR), measured in dB. The estimation methodology also provides error estimates for the measurements. The hypothesis is that the time history of the vertical drifts can be the foundation for irregularity forecasting. Irregularities are associated with strong increases in signal intensity.

In addition to the radar measurements, seasonal parameters, including the day of the year and the solar radiation flux at 10.7 cm, are used as features for machine learning.

### 2.2. Features and Labels

The data incorporated in the machine learning model consists of a set of features that describe or quantify each data sample along with a label possibly associated with those features. We are interested in a model that can receive these sample features and then predict the most likely label.

For this study, we use the average ion drift velocities measured by the ISR, the day of the year, and the solar flux as the features or inputs to the machine learning model. For each 5-min interval between 1200 and 1800 LT, the ion drift velocities are averaged over the 200–850 km altitude range using the squared inverse of the measurement errors as weights. The day of the year is represented by cartesian ( $x$ ,  $y$ ) coordinates in a unit circle to exploit the cyclical nature of this parameter. Finally, the solar flux for each measurement day is expressed in solar flux units as a separate feature. Together, there are 75 features for each day: 72 averaged vertical drifts, two DOY numbers, and one solar activity proxy.

The labels are inferred from the intensity of the backscattered radar signals. Irregularities associated with ESF are regarded as being present when the SNR of the coherent scatter exceeds a threshold. Space-time clusters of coherent scatter are identified in the data, and a label is derived from the maximum altitude of the cluster members.

**Table 1**  
*The Number of Measurements Was Categorized Based on Solar Flux and Time of Year*

Season	Low solar activity	High solar activity
Equinox	32	38
Winter Solstice	11	42
Summer Solstice	94	58

*Note.* The labels “Low Solar Activity” and “High Solar Activity” correspond to daily F10.7 values below and above 82 s.f.u. (solar flux units), respectively.

Plasma convective instabilities are usually observed shortly after sunset or thereafter, so we investigate the period between 1800 and 2300 LT. We define three labels: (a) a cluster is not observed, (b) a cluster is observed with a maximum altitude below 350 km, and (c) a cluster is observed with a maximum altitude above 350 km. Here, (b) corresponds roughly to bottom-side layers, which are common but less disruptive irregularity forms, whereas (c) corresponds roughly to topside radar plumes, which are less common but more disruptive (Hysell & Burcham, 1998; Woodman & La Hoz, 1976). Ideally, it would be desirable to predict the bottom side and topside scattering layers separately, but this will be a topic of future work.

### 2.3. Data Processing for Machine Learning

The data for this study were obtained from the CEDAR/Madrigal Database in HDF5 format and processed using the Python libraries `pandas`, `scikit-learn`, and `keras`. One of this study's main challenges is the marginal size of the data set. The ISR is run in campaign mode and does not produce a continuous data stream. Moreover, we limited the study to data acquired after 2010 when the particular experimental mode went into service. Campaign days that exceeded a specified threshold of missing data or presented excessive experimental variance in the ion drift velocity measurements were deemed unreliable and discarded, further reducing the size of the data set. Efforts were made to utilize samples that presented only a few missing data by interpolating from adjacent measurements. In total, 245 days were used to train the machine-learning models (Table 1).

It was important to build a robust, autonomous tool for detecting and quantifying clusters of coherent scatter since we expect this tool to continue to be used in the future on the ever-expanding JRO data set. This cluster identification is needed to filter isolated high SNR bins produced by experimental contamination or transients that will not be included in this analysis. The Python library `scikit-learn` has an implementation of the DBSCAN algorithm which was used to identify clusters of SNR bins above a predefined threshold. DBSCAN finds core samples of high density depending on two main parameters and extracts clusters from them. The parameters are `eps`, which defines a radius around an analyzed data point, and `min_samples`, the minimum number of samples inside a circle of radius `eps` around a data point to be considered part of the cluster (Schubert et al., 2017).

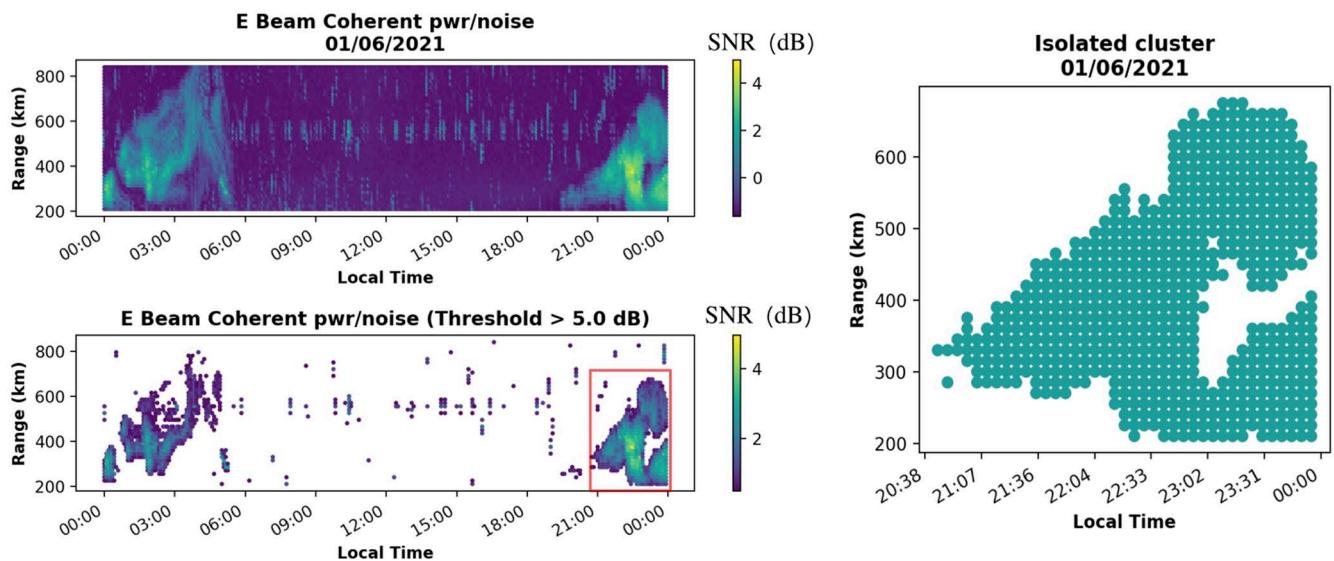
Several factors favored DBSCAN over other clustering algorithms. First, it performs well in applications with irregularly shaped clusters. Second, it extends clusters only in high-density areas of data points, effectively ignoring outliers in the data. The radio frequency environment is full of interference and clutter that is difficult to characterize, and this algorithm effectively ignores clusters or irrelevant data points. Before applying the DBSCAN algorithm, the data samples are filtered using a 5 dB threshold, discarding data points with insufficient SNR. However, some outliers are present after this filtering, and the clustering algorithm effectively mitigates this problem. This clustering technique provides useful information, including the maximum altitude of the ionospheric irregularities, their duration, the average intensity of the cluster, and the size of the cluster. Figure 1 illustrates the process of thresholding and cluster identification using DBSCAN for 1 day of data (Figure 2).

## 3. Model Details

We now describe the two machine learning models applied to our data. For this study, we used Random Forests (RF) to analyze the relationship between the features and the labels and extract feature importance and performance information. Additionally, a Convolutional Neural Network (CNN) was constructed and fitted to the data to improve the time-series classification accuracy further.

### 3.1. Label Classification Using Random Forests

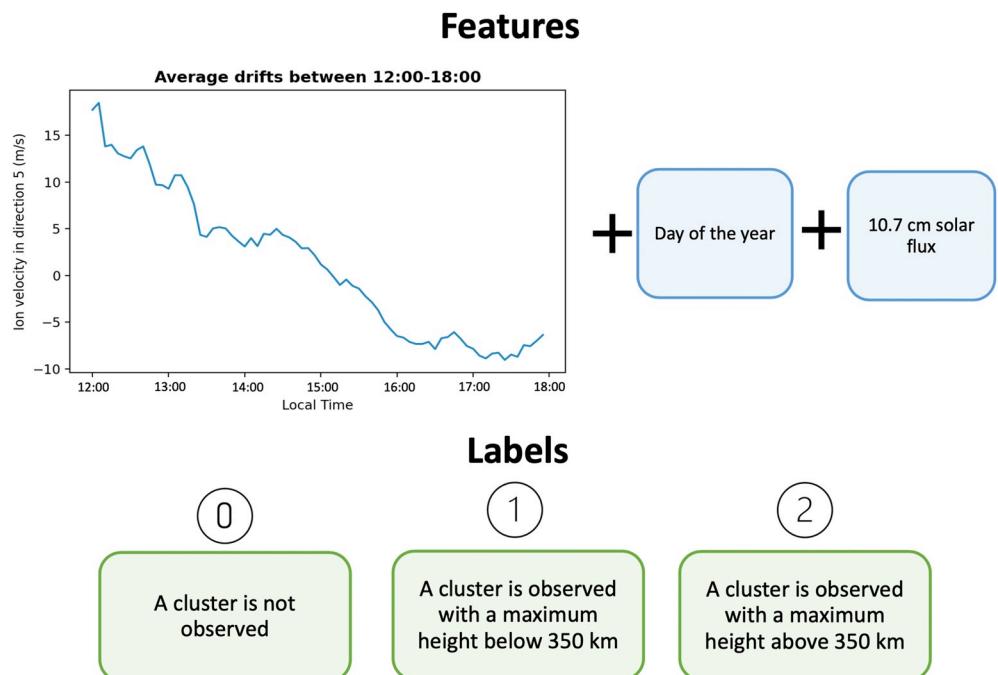
The first model applied to the data was Random Forests, an ensemble machine-learning technique that implements several individual decision trees for regression and classification problems (Breiman, 1996). The `scikit-learn` package was used to implement a supervised classification task on our data samples and extract individual feature importance that indicates which of the analyzed features is most relevant to each sample class. This classifier algorithm uses binary decision trees constructed using the Classification and



**Figure 1.** (Top left) SNR values versus altitude and local time. (Bottom Left) Only the bins above a threshold of 5 dB are retained. The red rectangle is filtered using the DBSCAN algorithm. (Right) DBSCAN-filtered SNR map derived from the previous plot.

Regression Trees algorithm. Each tree consists of several internal nodes and leaves representing features and attributes, respectively. In our data, the attributes correspond to the numeric values of those parametric features. The splitting decision for each node is based on a threshold value calculated to minimize the variance of the data in the two child nodes. In our model, the mean squared error is used to calculate the splitting threshold for each node, creating an iterative process that only stops when the tree can correctly classify all training data samples.

An individual decision tree presents high variance, leading to poor performance with unseen data. Random forests minimize this problem by taking random bootstrap samples (samples with replacement) from the training



**Figure 2.** Diagram illustrating the architecture of the model. The machine-learning model tries to reconstruct a function that maps average drifts, day of the year, and solar flux to the defined labels.

set and constructing several decision trees. In our classification task, the output of the RF classifier is the output mode of the individual trees.

The scikit-learn's package includes a built-in random forest feature importance module which computes the percentage relevance of each individual feature by analyzing how on average, it contributes to reducing the likelihood of a new unseen data sample being misclassified (Coppersmith et al., 1999). In our study, feature importance data provides insight into which portion of the time series better predicts the presence and location of SNR clusters after sundown.

### 3.2. Convolutional Neural Networks to Forecast Labels

In our study, a CNN was used as an alternative method of time-series classification (TSC), tuned in accordance with the results obtained from the RF model application. CNNs are commonly used in image classification tasks, where 2-dimensional arrays are filtered and analyzed in search of specific repeating patterns for samples belonging to the same output class. Recent studies have shown that this approach can be applied to a TSC problem and achieve a more accurate model than using recurrent neural networks (Fawaz et al., 2019). In our study, a 1-dimensional univariate array of points is convolved in search of specific patterns in average ion drift velocity variations over time.

Several motivations existed for using this model in our TSC problem. First, the batch normalization operation before each activation function accelerates the training of the network by reducing the covariance shift of the data, which cannot be implemented between time steps in networks like RNNs (Ioffe & Szegedy, 2015). Second, they perform relatively well where some samples are missing (Likowski et al., 2022). The samples obtained from the ISR often present missing data points, and despite applying interpolation techniques, the artificial generation of data points may pose a larger problem in other algorithms. Finally, aggregating a pooling layer after the convolution layers significantly reduces the computational cost when the network is deep (Liu et al., 2016).

In our study, we used a Fully Convolutional Neural Network (FCN) architecture (Wang et al., 2017) to classify the time-series sequence. The Python framework *Keras*, which is built atop *TensorFlow* 2, was used to implement the FCN. The designed architecture presents three convolutional layers, with 128, 256, and 128 filters, respectively, activated by a Rectified Linear Unit function; a 1-dimensional global average pooling layer, which emphasizes the portions of the time series that most contributed to the predicted class; and a dense final layer activated using a softmax function that normalizes the final output to get a probability estimate. The best parameters for batch size and the number of epochs were found through an exhaustive search performed by the *GridSearchCV* method for model selection implemented in *Keras*. Finally, initial tests showed overfitting, so dropout layers were inserted between layers as regularization.

## 4. Results and Discussion

A total of 275 days of measurements made between 2010 and 2021 were available to train and test our machine-learning models. For our analyses, the data were randomly split into a training subset containing 245 samples, and the remaining 30 days were used as validation data to calculate the accuracy of the model.

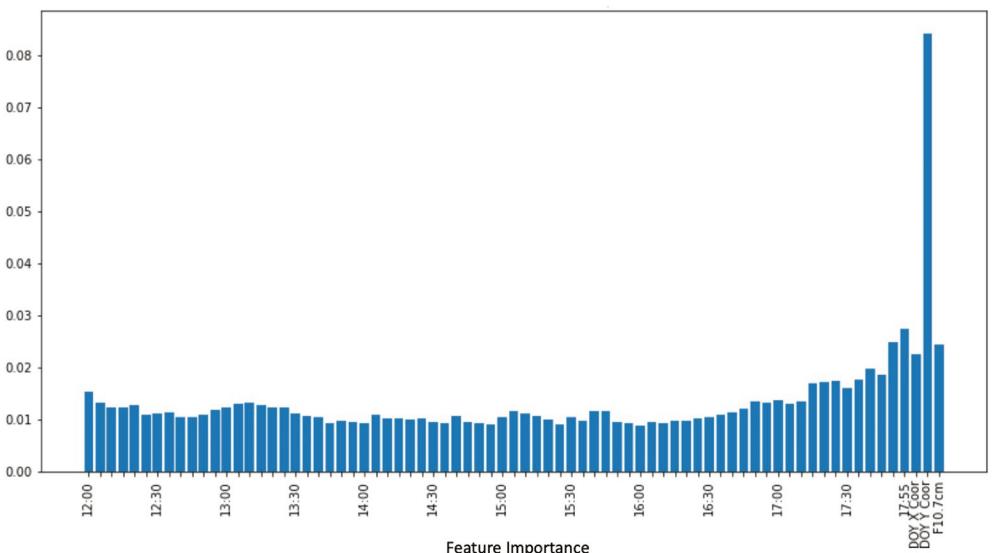
Given the limited availability of days of measurements for this study, the performance results were obtained by taking the average of several model applications. In our RF classification, the performance was calculated with the *accuracy\_score* metric in the Scikit-learn Python package. This function computes subset accuracy for multi-label classification by comparing the predicted and true testing labels, which must match exactly. The model presented a high variability in testing accuracy. This high variance can be explained by the low number of samples available for training. To mitigate this problem, we took the average of 500 random forests, with 100 estimators in each forest. The averaged confusion matrix is presented in Table 2, and it shows the average number of samples that were predicted for each label versus the true number of samples for each class. This provides a visual representation of the overall performance of the random forest algorithm. The rows represent the actual labels derived from analyzing the testing data, while the columns show the labels forecasted by our model. For instance, our model's average prediction for label 1 was actually label 0, occurring 3 times. Moreover, the model accurately predicted labels 0, 1, and 2, with frequencies of 10, 4.7, and 34 times, respectively.

**Table 2**  
*Average Confusion Matrix for RF Classification*

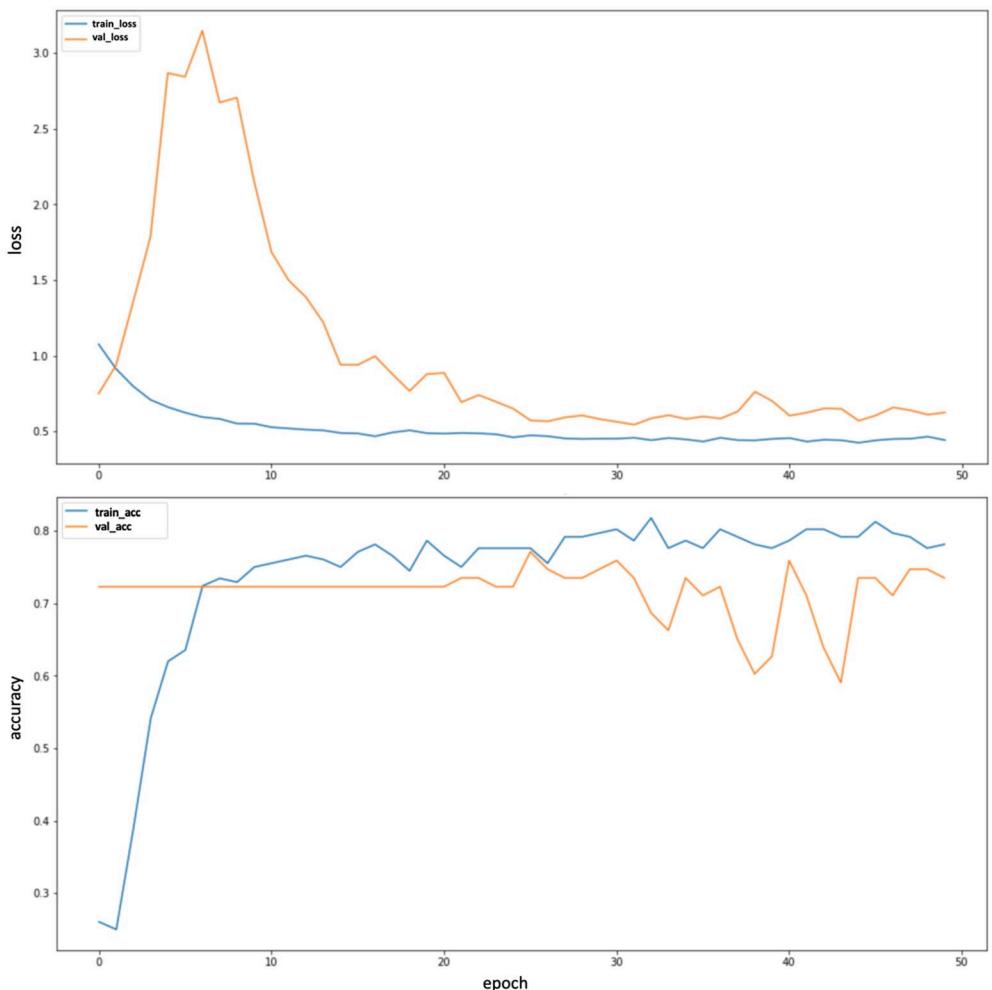
True Label	Predicted Label		
	0	1	2
0	10	3	2.8
1	2	4.7	5.5
2	8.8	12	34

As observed, the model successfully identifies samples exhibiting no clusters and high-altitude clusters. However, the model struggles to find a good feature-label relationship for low-altitude clusters. This indicates that the existing patterns in the time series are not incisive enough to differentiate clusters of varying altitudes in the RF classification. Nonetheless, there exists a clear relationship between the features and the presence of high SNR clusters.

The individual feature importance for our model (Figure 3) indicates that seasonal parameters are strongly correlated to the presence of clusters. This feature is the most relevant during the classification task as expected (Fejer et al., 1979). The parameter that contributes the most to the predicted class is the day of the year, represented as  $(x, y)$  coordinates of a unit circle. It can be observed that the  $y$ -coordinate is significantly more relevant for the model than the  $x$ -coordinate, which can be explained by one of the following hypotheses: (a) in the case of correlated



**Figure 3.** RF Feature Importance for classification.



**Figure 4.** Evolution of the loss function and model accuracy for different stages of the training.

features, the built-in feature importance method can select one of the features and neglect the importance of the other (Breiman, 2001); (b) the presence of convective plasma instabilities is directly correlated to the sinusoidal component of the time series. Furthermore, the importance of the averaged drift time series is most pronounced right before 1800 LT, which indicates that the model may obtain similar results by shortening the time series to the last few hours.

The results obtained with the FCN are summarized in Figure 4. Both the training and validation loss functions decrease for increasing epochs. Even though both curves seem to converge, we see that the training loss function tends to a value half of the validation loss function. This could be a combination of two factors. First, the validation data set does not represent the training data, so the model will always have difficulty modeling new data. Second, the input parameters of the model may not have enough predictive power to capture the clustering labels compared to their variability. Nevertheless, the accuracy oscillates around 75%, which is significantly greater than the 33% of random choice. It is important to emphasize that this latter claim does not imply that we expect our model to significantly outperform simpler statistical models or a similar model that does not consider the information of the drifts. Increasing the number of training epochs would produce overfitting.

## 5. Summary and Conclusions

In this work, we built two models based on machine learning to predict the occurrence of ionospheric irregularities associated with ESF based on vertical drift data measured at the JRO. The occurrence of these irregularities is associated with clusters of high SNR and is identified using the DBSCAN clustering algorithm. The processed radar data are then labeled with 0 (no cluster), 1 (cluster located below 350 km), and 2 (cluster located above

350 km). Then, we used Random Forests and Convolutional Neural Networks to build models that map averaged vertical drifts, day of the year, and solar activity to these clustering labels.

Our results suggest that even though climatology is still the strongest predictor, the vertical drifts from a few hours before sundown have predictive power. Despite the small data sets, the Neural Network achieves an accuracy of around 75%, which is remarkable considering the simplicity of the model. In the future, new data sets may be assimilated to increase the predictive power of the model.

## Data Availability Statement

The ISR velocity drift data used for the training of our model are available through the Madrigal database (Scipion, 2018).

## Acknowledgments

This project is supported by NSF award no. AGS-2028032 to Cornell University. The Jicamarca Radio Observatory is a facility of the Instituto Geofísico del Perú operated with support from NSF award AGS-2213849 through Cornell. The help of the staff is much appreciated.

## References

- Booker, H. G., & Wells, H. W. (1938). Scattering of radio waves by the *F* region. *Terrestrial Magnetism and Atmospheric Electricity*, 43(3), 249–256. <https://doi.org/10.1029/te043i003p00249>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/bf00058655>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chapagain, N. P., Fejer, B. G., & Chau, J. L. (2009). Climatology of postsunset equatorial spread *F* over Jicamarca. *Journal of Geophysical Research*, 114(A7), A07307. <https://doi.org/10.1029/2008JA013911>
- Chen, C. H., Lin, C., Chen, W. H., & Matsuo, T. (2017). Modeling the ionospheric prereversal enhancement by using coupled thermosphere-ionosphere data assimilation. *Geophysical Research Letters*, 44(4), 1652–1659. <https://doi.org/10.1002/2016GL071812>
- Coppersmith, D., Hong, S. E., & Hosking, J. R. (1999). Partitioning nominal attributes in decision trees. *Data Mining and Knowledge Discovery*, 3(2), 197–217. <https://doi.org/10.1023/A:1009869804967>
- Eccles, J. V., Maurice, J. P. S., & Schunk, R. W. (2015). Mechanisms underlying the prereversal enhancement of the vertical plasma drift in the low-latitude ionosphere. *Journal of Geophysical Research*, 120(6), 4950–4970. <https://doi.org/10.1002/2014JA020664>
- Farley, D. T., Balsley, B. B., Woodman, R. F., & McClure, J. P. (1970). Equatorial spread *F*: Implications of VHF radar observations. *Journal of Geophysical Research*, 75(34), 7199–7216. <https://doi.org/10.1029/ja075i034p07199>
- Farley, D. T., Bonelli, E., Fejer, B. G., & Larsen, M. F. (1986). The prereversal enhancement of the zonal electric field in the equatorial ionosphere. *Journal of Geophysical Research*, 91(A12), 13723–13728. <https://doi.org/10.1029/ja091ia12p13723>
- Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, 33(4), 917–963. <https://doi.org/10.1007/s10618-019-00619-1>
- Fejer, B. G., Farley, D. T., Woodman, R. F., & Calderon, C. (1979). Dependence of equatorial *F*-region vertical drifts on season and solar cycle. *Journal of Geophysical Research*, 84(A10), 5792–5796. <https://doi.org/10.1029/ja084ia10p05792>
- Fejer, B. G., & Kelley, M. C. (1980). Ionospheric irregularities. *Reviews of Geophysics*, 18(2), 401–454. <https://doi.org/10.1029/rg018i002p00401>
- Gentile, L. C., Burke, W. J., & Rich, R. G. (2006). A global climatology for equatorial plasma bubbles in the topside ionosphere. *Annales Geophysicae*, 24(1), 163–172. <https://doi.org/10.5194/angeo-24-163-2006>
- Hysell, D. L., & Burcham, J. (1998). JULIA radar studies of equatorial spread *F*. *Journal of Geophysical Research*, 103(A12), 29155–29167. <https://doi.org/10.1029/98ja02655>
- Hysell, D. L., Fang, T. W., & Fuller-Rowell, T. J. (2022). Modeling equatorial *F*-region ionospheric instability using a regional irregularity model and WAM-IPE. *Journal of Geophysical Research*.
- Ioffe, S., & Szegedy, C. (2015). In *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. (Vol. 1).
- Kelley, M. C., Makela, J. J., de la Beaujardiere, O., & Retterer, J. (2011). Convective ionospheric storms: A review. *Reviews of Geophysics*, 49(2), RG2003. <https://doi.org/10.1029/2010RG000340>
- Kil, H. (2015). The morphology of equatorial plasma bubbles—A review. *Journal of Astronomy and Space Sciences*, 32(1), 13–19. <https://doi.org/10.5140/JASS.2015.32.1.13>
- Kil, H., Paxton, L. J., & Oh, S. J. (2009). Global bubble distribution seen from ROCSAT-1 and its association with the evening prereversal enhancement. *Journal of Geophysical Research*, 114(A6), A06307. <https://doi.org/10.1029/2008JA013672>
- Kudeki, E., Bhattacharyya, S., & Woodman, R. F. (1999). A new approach in incoherent scatter *F* region E x B drift measurements at Jicamarca. *Journal of Geophysical Research*, 104(A12), 28145–28162. <https://doi.org/10.1029/1998ja00110>
- Likowski, M. P., Smieja, M., Struski, L., & Tabor, J. (2022). Misconv: Convolutional neural networks for missing data. <https://doi.org/10.1109/WACV51458.2022.00297>
- Liu, S., Qi, X., Shi, J., Zhang, H., & Jia, J. (2016). In *Multi-scale patch aggregation (MPA) for simultaneous detection and segmentation*. <https://doi.org/10.1109/CVPR.2016.342>
- Richmond, A. D., Fang, T. W., & Maute, A. (2015). Electrodynamics of the equatorial evening ionosphere: 1. Importance of winds in different regions. *Journal of Geophysical Research*, 120(3), 2118–2132. <https://doi.org/10.1002/2014JA020934>
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, 42(3), 1–21. <https://doi.org/10.1145/3068335>
- Scipion, J. R. O. D. (2018). Data from the cedar madrigal database. Retrieved from <http://cedar.openmadrigal.org/getCitationGroup?id=1014>
- Wang, Z., Yan, W., & Oates, T. (2017). In *Time series classification from scratch with deep neural networks: A strong baseline*. <https://doi.org/10.1109/IJCNN.2017.7966039>
- Woodman, R. F. (1970). Vertical drift velocities and east-west electric fields at the magnetic equator. *Journal of Geophysical Research*, 75(31), 6249–6259. <https://doi.org/10.1029/ja075i031p06249>
- Woodman, R. F. (2009). Spread *F*- an old equatorial aeronomy problem finally resolved? *Annales Geophysicae*, 27(5), 1915–1934. <https://doi.org/10.5194/angeo-27-1915-2009>
- Woodman, R. F., & La Hoz, C. (1976). Radar observations of *F* region equatorial irregularities. *Journal of Geophysical Research*, 81(31), 5447–5466. <https://doi.org/10.1029/ja081i031p05447>