Forecasting Spread F at Jicamarca

Reynaldo O. Rojas[1], Enrique L. Rojas[2], Jhassmin A. Aricoché[3], Marco A. Milla[4]
1. Universidad de Ingeniería y Tecnología, Lima, Perú
2. Cornell University, New York, USA
3. Jicamarca Radio Observatory, Lima, Perú
4. Sección Electricidad y Electrónica, Pontificia Universidad Católica del Perú, Lima, Perú
reynaldo.rojas@utec.edu.pe

Introduction

What is Spread F?

Spread F is a phenomenon that occurs in the F layer of the Ionosphere and is characterized by plasma depletions. It usually starts at early times of the night.

Why does it matter?

It can have a negative impact on radio communication systems and because of this, it is of interest to develop a model that can predict its occurrence.

How can we predict its occurrence?

Radars like digisondes and JULIA (Jicamarca Unattended Long-term Investigations of the Ionosphere and Atmosphere) have observed the Ionosphere at Jicamarca for decades. These measurements along with geophysical parameters were harnessed to train a Machine Learning model that makes a daily prediction about its occurrence. Although our model has only been validated with Jicamarca's dataset, it may be used for other longitudes. Furthermore, since the only local measurements used during training were Spread F occurrences and the virtual height of the F layer, the retraining process can easily be done on a single station with an ionosonde receiver.

Model

Architecture

It is a Multilayer Perceptron with ELU as activation function. It outputs a real number has an image that can be interpreted as the probability of occurrence when passed by a sigmoid function.

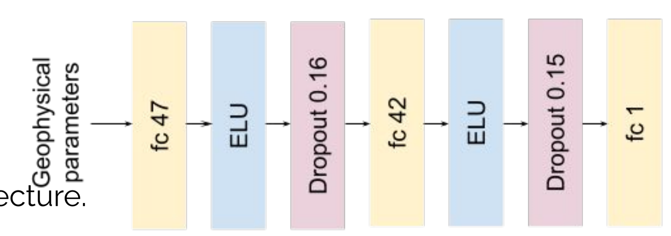


Figure 1. Neural Network Architecture.

In figure 1, fc X is a fully-connected layer with X units. An activation function adds non-linearity to the network allowing it to learn from higher-level representations. Dropout p is a layer which, with probability p, sets to 0 the output of any unit.

Optimization

Training

Parameter optimization was carried out with the Adam algorithm for 30 epochs. The loss function used was `torch.nn.BCEWithLogitsLoss`.

Hyper-parameter configuration search

The model proposed corresponds to the best Optuna trial. This library implements the Sequential Model-Based Optimization algorithm that uses a Tree-structured Parzen Estimator as surrogate. As shown below, we conducted 800 trials.

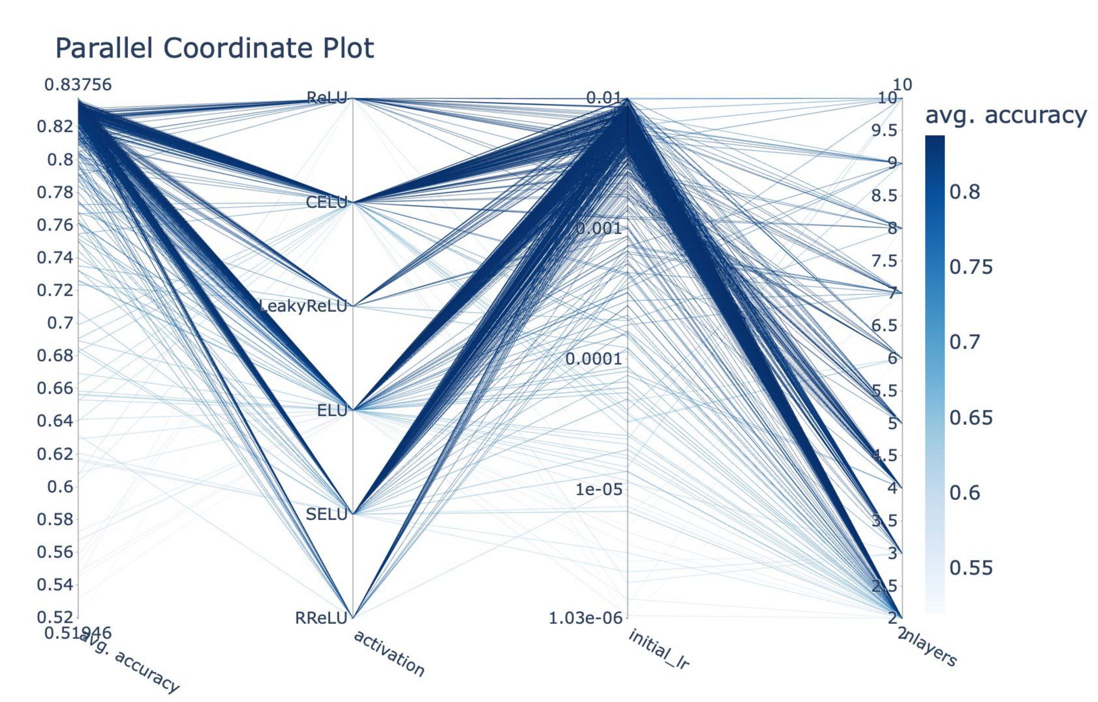


Figure 2. Hyper-p

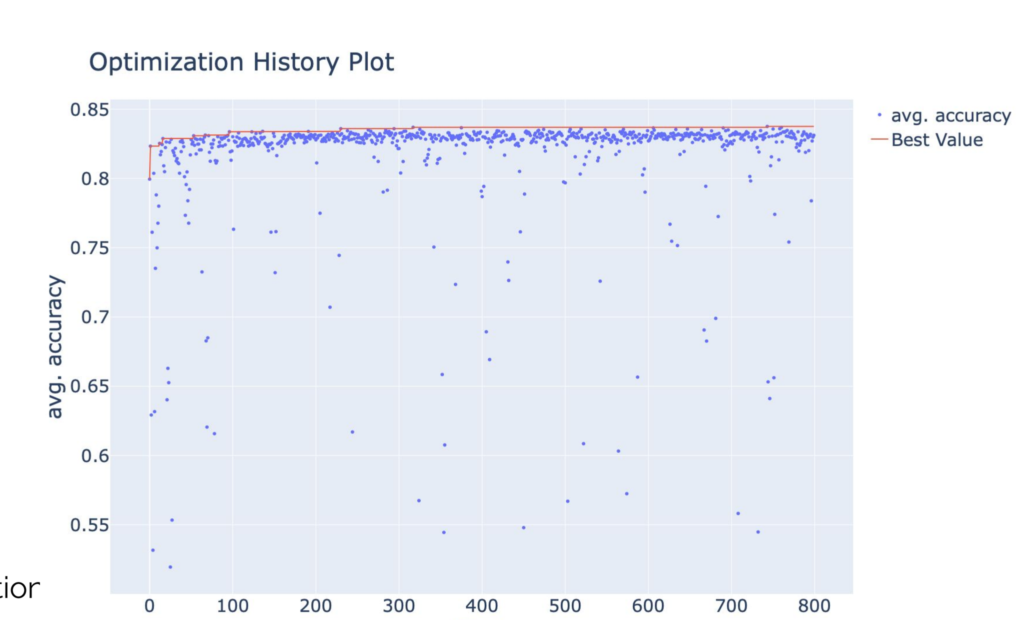
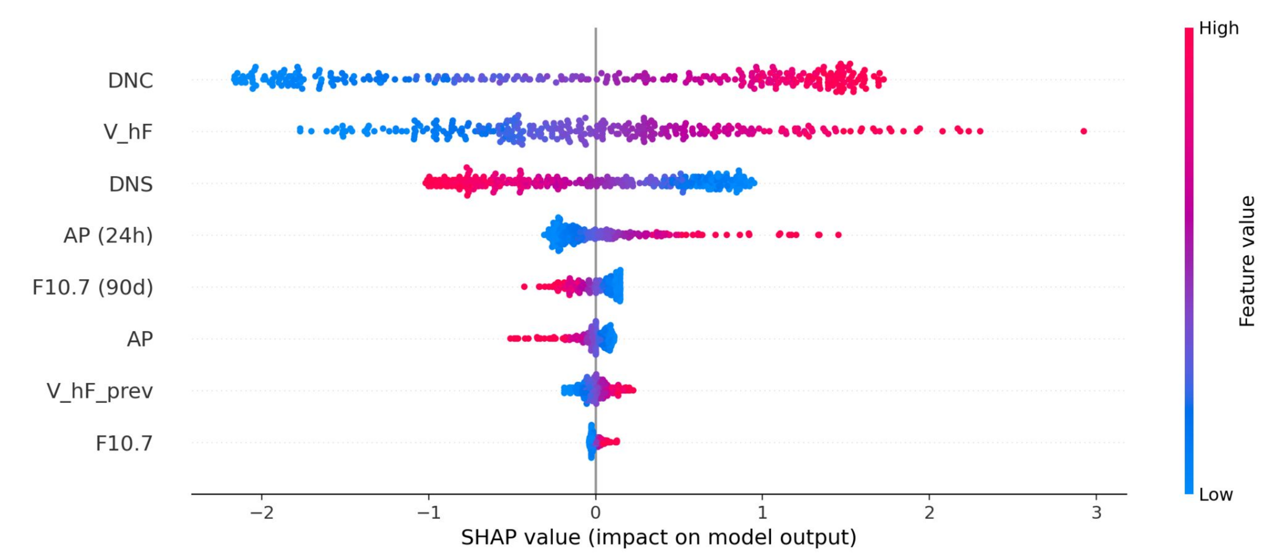


Figure 3. Optimizatior

Input Sensitivity

Figure 4 presents feature importance and the effect that each feature on the predictions made by the model for 300 instances from the dataset. Features are sorted in descending order according their importance.

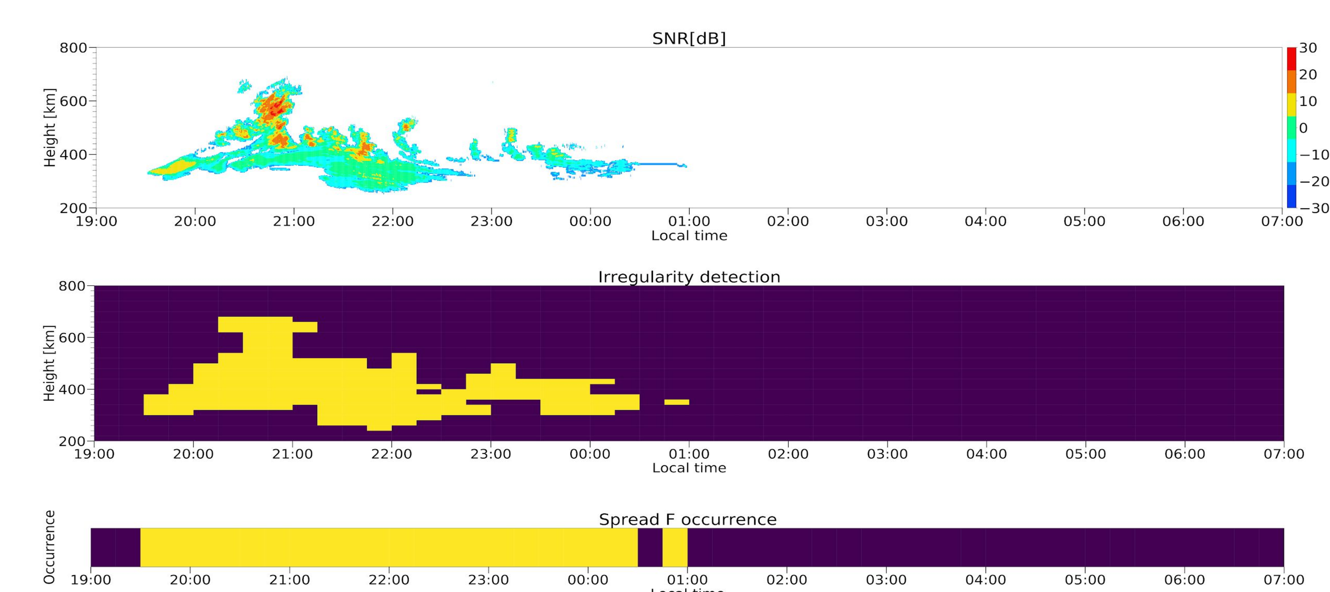


Measurements

JULIA (Jicamarca Unattended Long-term Investigations of the Ionosphere and Atmosphere) radar observations were used to determine ESF occurrences. The geophysical parameters used as inputs consist of Jicamarca's digisonde measurement as well as global parameters. The dataset used spans the years 2000 - 2020.

Data processing

We did the same ESF characterization process as in Zhan, Rodrigues and Milla (2018). Since all of the relevant information for our supervised learning algorithm lies on the time axis, we collapsed the height axis.



ESF occurrences for each day were stored as a time series and later merged with digisonde measurements and global parameters.

Input pre-processing

All inputs passed to the input must be between 0 and 1 to make it easier for the model to learn a representation and mapping. Geophysical parameters take a range of values that escape this small range. We used `sklearn.preprocessing.MinMaxScaler` to fit the values into the desired range. The day of the year however, is a periodic variable and, as such, should not be dependent on the choice of origin (Bishop, 2011).

$$DNS = \sin(2\pi D/365),\ DNC = \cos(2\pi D/365),\ D\text{: Day of year (1-365)}$$

Datasets

We split our entire dataset in 3 subsets: Training, Validation and Testing datasets.

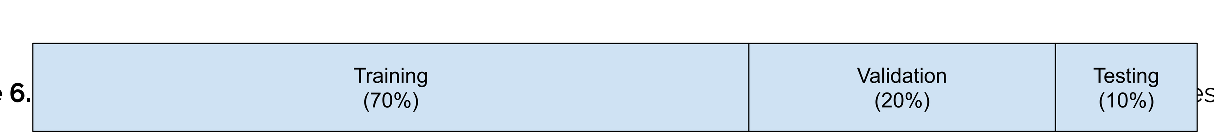


Figure 6. ...esting.

The training dataset was later partitioned into folds and the model was trained from scratch with each of them. Our goal was to find a hyper-parameter configuration while optimizing the average accuracy across the folds..

Results

These results correspond to the evaluation and comparison of our model and FIRST on the testing dataset. It is important to point out that, in this work, FIRST is evaluated with occurrences obtained from the characterization presented earlier as opposed to the evaluation in Anderson and Redmon (2017), which apparently used occurrences from manually labeled ionograms. There were a total of 69 days for which FIRST did not make a prediction.

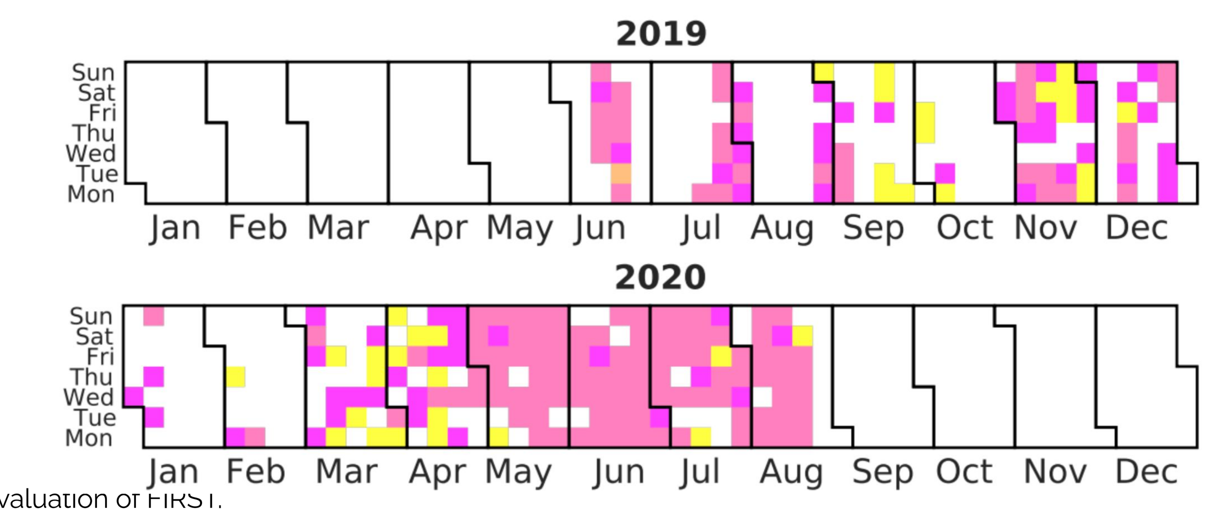


Figure 7. Evaluation of FIRST.

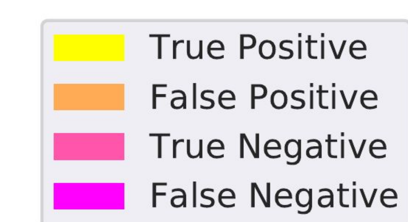


Figure 8. Color map for the confusion calendars and matrices.
True positive: The model predicted that Spread F would occur, and it did.
False positive: The model predicted that Spread F would occur, but it did not.
True negative: The model predicted that Spread F would not occur, and it did not.
False negative: The model predicted that Spread F would not occur, but it did.

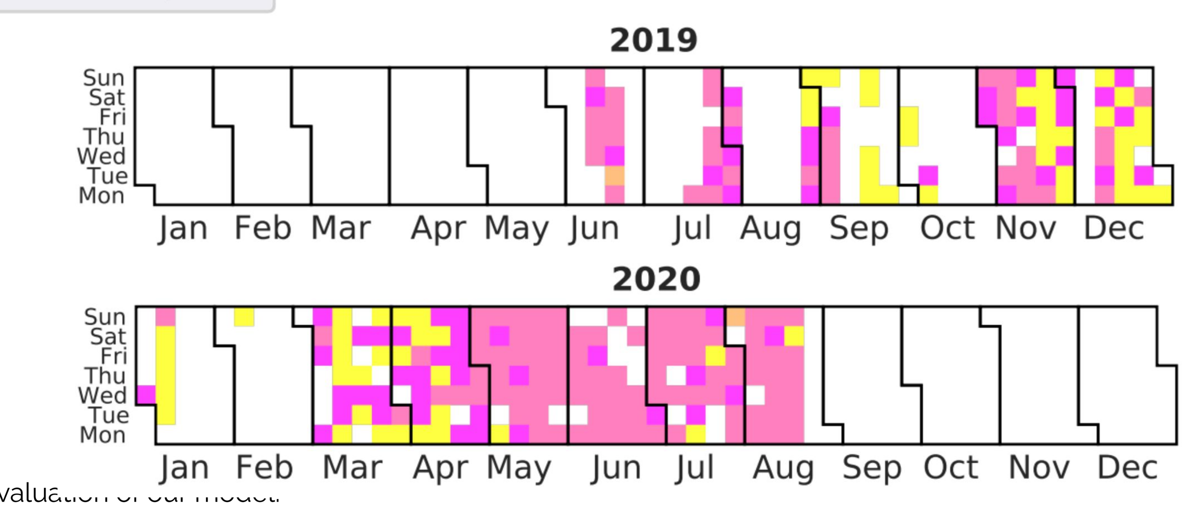


Figure 9. Evalua...

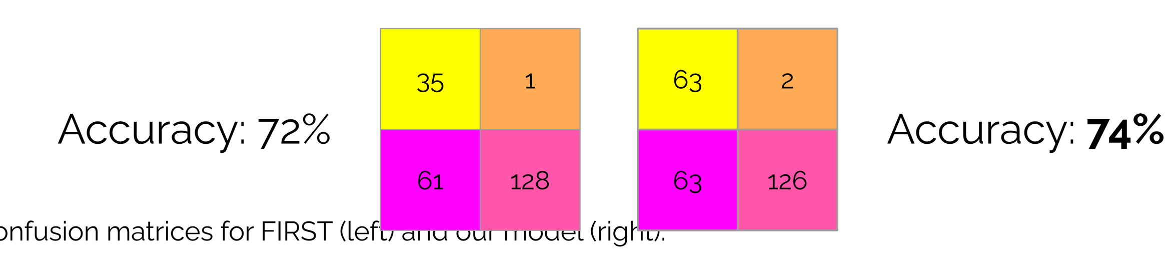


Figure 10. Confusion matrices for FIRST (left and our model (right).

Conclusions

The most important geophysical parameters appear to be the day of the year and h'F. In addition, our preliminary results suggest that the predictive power of our model is slightly better than FIRST but further analysis is required to validate this claim. Previously, we conducted the training procedure with a bigger architecture and without Optuna, K-Fold Cross Validation or early-stopping, obtaining a higher accuracy for the testing dataset. We hypothesize that this happens because the patience of the early-stopping procedure is too low and the models that can achieve higher accuracies need more epochs or perhaps because the average accuracy is not the best metric to optimize and we should use a weighted average instead.

Acknowledgements

Special thanks to Alejandro Palacios for the assistance provided with the datasets. We also thank the support with the Jicamarca Radio Observatory staff. The Jicamarca Radio Observatory is a facility of the Instituto Geofisico del Peru operated with support from the NSF AGS-1732209 through Cornell University.

References

Anderson, D. N., and R. J. Redmon (2017), Forecasting scintillation activity and equatorial spread F, *Space Weather, 15,* 495–502, doi:10.1002/2016SW001554.
Bishop, C. M. (2006), *Pattern Recognition and Machine Learning (Information Science and Statistics),* Springer.
Chapagain, N. P., B. G. Fejer, and J. L. Chau (2009), Climatology of postsunset equatorial spread F over Jicamarca, *J. Geophys. Res., 114,* A07307, doi:10.1029/2008JA013911.
Fejer, B. G., Scherliess, L., and de Paula, E. R. (1999), Effects of the vertical plasma drift velocity on the generation and evolution of equatorial spread *F, J. Geophys. Res.*, 104(A9), 19859– 19869, doi:10.1029/1999JA900271.
Hysell, D. L., & Burcham, J. D. (2002). Long term studies of equatorial spread F using the JULIA radar at Jicamarca. *Journal of Atmospheric and Solar-Terrestrial Physics, 64* (12-14), 1531-1543. https://doi.org/10.1016/S1364-6826(02)00091-3
Kelley, M. C. (2009). *The earth's ionosphere: Plasma physics and electrodynamics* (2nd ed.). Academic Press.
Paszke, A. et al. (2019) "PyTorch: An Imperative Style, High-Performance Deep Learning Library", in Wallach, H. et al. (reds) *Advances in Neural Information Processing Systems 32.* Curran Associates, Inc., bll 8024–8035. Available at: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
Pedregosa, F. et al. (2011) "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research,* 12, bll 2825–2830.
Redmon, R. J., D. Anderson, R. Caton, and T. Bullett (2010), A Forecasting Ionospheric Real-time Scintillation Tool (FIRST), *Space Weather, 8,* S12003, doi:10.1029/2010SW000582
Zhan, W., Rodrigues, F., & Milla, M. (2018). On the genesis of postmidnight equatorial spread F: Results for the American/Peruvian sector. *Geophysical Research Letters, 45,* 7354–7361. https://doi.org/10.1029/2018GL078822
Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). Association for Computing Machinery, New York, NY, USA, 2623–2631. DOI:https://doi.org/10.1145/3292500.3330701
Lundberg, S. M. en Lee, S.-I. (2017) "A Unified Approach to Interpreting Model Predictions", in Guyon, I. et al. (reds) *Advances in Neural Information Processing Systems.* Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf