

# A Machine Learning Approach to Predicting Earth's Open-Close Boundary

Arnav Singh<sup>1</sup>

<sup>1</sup>Dartmouth College

## Key Points:

- Machine learning approaches (regression and classification) achieve robust OCB predictions from solar wind and geomagnetic indices.
- The 5-minute AE index and MLT consistently rank as top features driving boundary variability.
- Hemispheric data gaps and nightside variability highlight the need for balanced satellite coverage and enhanced storm-time modeling.
- Combining probabilistic (classification) and continuous (regression) methods offers both uncertainty estimates and precise boundary positioning.

---

Corresponding author: Arnav Singh, [arnav.singh.26@dartmouth.edu](mailto:arnav.singh.26@dartmouth.edu)

14 **Abstract**

15 The Open-Closed Boundary (OCB) in Earth’s magnetosphere represents a crucial de-  
 16 marcation between closed magnetic field lines and those connected to the interplanetary  
 17 magnetic field. We present a novel machine learning approach to predict OCB locations  
 18 using multi-satellite observations and solar wind parameters. Our work analyzes high-  
 19 resolution (5-minute) data from multiple sources including OMNI and satellite obser-  
 20 vations, focusing on the relationship between the OCB and various magnetospheric in-  
 21 dices. Using a comprehensive machine learning approach incorporating both regression  
 22 and classification techniques, we achieve up to 80% accuracy. Additionally, we can show  
 23 that models can be made to predict OCB latitude across magnetic local times (MLT),  
 24 using Neural Networks. Regression models demonstrate moderate success with RMSE  
 25 values around 3 degrees latitude, while classification approaches show compromising lat-  
 26 itude resolution can yield significant increases in accuracy. Our analysis reveals system-  
 27 atic variations in prediction accuracy between dayside and nightside regions, with dis-  
 28 tinct feature importance patterns in each sector. We identify the AE index and MLT as  
 29 critical predictors, with their relative importance varying between hemispheres and lo-  
 30 cal time sectors. The ionosphere’s b2i boundary adds crucial data for nightside models.  
 31 Hemispheric asymmetries in data availability provide valuable insights for future satel-  
 32 lite mission planning. This research advances our understanding of magnetosphere-ionosphere  
 33 coupling dynamics and provides a foundation for improved space weather prediction ca-  
 34 pabilities.

35 **Plain Text Summary**

36 Earth’s magnetosphere protects our planet from the constant flow of charged par-  
 37 ticles streaming from the Sun. One important region in this protective shield is the Open-  
 38 Closed Boundary (OCB), which marks where magnetic field lines go from being “closed”  
 39 (looping back to Earth) to “open” (connecting out into space). The location of this bound-  
 40 ary changes with space weather conditions and helps scientists understand how energy  
 41 from the Sun travels into Earth’s upper atmosphere.

42 In this study, we use data from orbiting satellites and solar wind measurements to  
 43 predict the OCB’s position. Traditional approaches have often tried to treat the bound-  
 44 ary as a continuous line, which we also did, but we found that treating it in simpler, “cat-  
 45 egorical” terms—like “higher latitude” or “lower latitude”—can actually boost predic-  
 46 tion accuracy to around 80%. This means our machine learning models were correct four  
 47 out of five times in telling which range of latitudes the boundary would fall into.

48 We also looked at which pieces of information mattered the most. For the daytime  
 49 side of the Earth, the AE index—a measure of geomagnetic activity—proved most im-  
 50 portant, while on the night side, the local time information mattered more. Addition-  
 51 ally, differences in satellite coverage across the Northern and Southern Hemispheres af-  
 52 fected how well the models performed, a finding that highlights the importance of com-  
 53 prehensive satellite missions. Overall, our results suggest that classifying the boundary’s  
 54 location into simpler categories can help make more consistent, reliable predictions, im-  
 55 proving our understanding of how the magnetosphere responds to solar wind changes.

56 **1 Introduction**

57 Understanding the dynamics of Earth’s magnetosphere-ionosphere system requires  
 58 accurate determination of the Open-Closed magnetic field line Boundary (OCB). This  
 59 boundary serves as a critical marker for solar wind-magnetosphere coupling and plays  
 60 a fundamental role in space weather phenomena (Newell et al., 2009). The OCB repre-  
 61 sents the demarcation between closed magnetic field lines, which have both footprints  
 62 in the ionosphere, and open field lines that connect to the interplanetary magnetic field

(IMF). This boundary's location and movement provide crucial information about energy transfer processes in the magnetosphere (Milan et al., 2015).

Despite its importance, consistent determination of the OCB location remains challenging due to the complex interplay of solar wind parameters, geomagnetic activity, and magnetospheric configuration (Lockwood et al., 2016). Traditional methods of OCB identification rely heavily on individual satellite observations, leading to spatial and temporal gaps in coverage. These gaps significantly limit our ability to continuously monitor magnetospheric dynamics and predict space weather events.

Recent advances in machine learning techniques offer promising approaches to combine multi-satellite data and predict OCB locations with improved accuracy (Camporeale, 2019). These methods require careful validation and understanding of the underlying physical parameters that influence boundary dynamics. The Newell coupling parameter (Newell et al., 2007), which we incorporate in our analysis, has proven particularly helpful for understanding solar wind-magnetosphere coupling.

Our work addresses these challenges through a comprehensive machine learning approach, combining regression and classification techniques to predict OCB location using multiple data sources. We focus on several key objectives:

1. Development of accurate prediction models using high-resolution magnetospheric data
2. Investigation of the relative importance of different physical parameters in controlling boundary location
3. Predicting OCB location across all MLT's for a given conditions
4. Analysis of hemispheric asymmetries in boundary dynamics
5. Assessment of machine learning techniques for operational space weather applications

## 2 Data and Methods

### 2.1 Data Sources and Preprocessing

Our study utilizes high-resolution (5-minute) data from multiple sources, primarily accessed through NASA's Space Physics Data Facility (SPDF). The core dataset comprises approximately 2 million initial observations from the OMNI database from 1983-2012, which provides thoroughly cross-calibrated and time-shifted solar wind parameters at Earth's bow shock nose. These parameters include magnetic field components, plasma parameters, and derived quantities essential for understanding solar wind-magnetosphere coupling.

The preprocessing pipeline implemented several critical steps to ensure data quality and temporal alignment. First, we performed temporal matching between OCB observations and OMNI data using a 5-minute tolerance window. This matching process accounted for the inherent uncertainties in timing relationships between solar wind parameters and magnetospheric responses. Data points lacking proper temporal matching were systematically removed to maintain data integrity.

Satellite coverage introduced significant constraints on data availability. The Defense Meteorological Satellite Program (DMSP) satellites, with their 101-minute orbital periods, provided our primary OCB observations, computed by Tom Sotirelis. This orbital characteristic resulted in systematic gaps in spatial coverage and created notable differences in data availability between hemispheres and MLT: 170,000 northern hemisphere samples versus 300,000 southern hemisphere samples for nightside observations, and bimodal histogram for MLT seen later in Figure 8.

110            **2.2 Feature Selection**

111            Feature selection followed a rigorous process that combined correlation analysis with  
 112 physical understanding of magnetospheric dynamics. Initial correlation analysis in Fig-  
 113 ure 1 revealed significant relationships between various solar wind and magnetospheric  
 114 parameters. We identified and addressed multicollinearity issues, particularly among ge-  
 115 omagnetic indices, to ensure model stability and interpretability. However, slight differ-  
 116 ences in the indices were valuable in training the different models. So we opted to not  
 117 discard all of some, but rather let the model figure it out.



**Figure 1.** Preliminary Correlation Matrix showcasing the various  $R^2$  values between the features.

118            The final feature set included:

1. Magnetic Local Time (OCB\_mlt or b6\_mlt): Essential for capturing diurnal variations in boundary location
2. Magnetic field parameters (MAG\_AVG\_B-VECTOR\_nT, BZ\_GSM\_nT): Critical for understanding magnetic field topology
3. Solar wind parameters (FLOW\_SPEED\_GSE\_km/s, PROTON\_DENSITY\_ncc): Primary drivers of magnetospheric dynamics
4. Flow pressure (FLOW\_PRESSURE\_nPa): Key parameter for magnetopause location
5. Geomagnetic indices (5-M\_AE\_nT, SYMH\_INDEX\_nT, 5-M\_PC(N)-INDEX\_): Indicators of global magnetospheric activity
6. Newell coupling parameter derivative (dFMP\_dt): Measure of solar wind-magnetosphere coupling efficiency
7. B2i magnetic latitude (b2i\_mag\_lat, nightside only): Additional boundary reference for nightside predictions

133            Feature scaling employed standardization (zero mean, unit variance) to ensure con-  
 134 sistent treatment across different parameter ranges. Missing values, particularly in so-  
 135 lar wind parameters during data gaps were removed to ensure data integrity.

136            **2.3 Machine Learning Implementation**

137     This paper employs both regression and classification approaches to predict OCB  
 138     latitude. Initially, we explored a comprehensive suite of regression models to capture the  
 139     continuous nature of the boundary location. The regression models included traditional  
 140     linear methods with varying regularization techniques: Linear Regression provided a base-  
 141     line approach assuming direct linear relationships between features, while Ridge Regres-  
 142     sion incorporated L2 regularization to prevent overfitting by penalizing large coefficients  
 143     (Hoerl & Kennard, 1970). Lasso Regression, utilizing L1 regularization, performed im-  
 144     plicit feature selection by driving some coefficients to zero (Tibshirani, 1996), and Elas-  
 145     ticNet combined both L1 and L2 penalties to balance between feature selection and co-  
 146     efficient stability.

147     For handling non-linear relationships, we implemented ensemble methods includ-  
 148     ing Random Forest Regression, which combines multiple decision trees through bootstrap  
 149     aggregation (bagging) to reduce overfitting (Breiman, 2001). Gradient Boosting and its  
 150     optimized implementation XGBoost provided sequential tree building focused on poorly  
 151     predicted samples (Chen & Guestrin, 2016). Additionally, we developed a neural net-  
 152     work regression model using a Multi-layer Perceptron architecture with three hidden lay-  
 153     ers (256, 128, and 64 neurons) employing ReLU activation functions and the Adam op-  
 154     timizer with learning rate scheduling (Goodfellow et al., 2016).

155     These regression models achieved promising results, with Root Mean Square Er-  
 156     rror (RMSE) values around 3 degrees latitude and Mean Absolute Error (MAE) under  
 157     3 degrees. The Random Forest regressor demonstrated particularly strong performance,  
 158     achieving an  $R^2$  of 0.449 for dayside predictions and a  $R^2$  of 0.49 for nightside predic-  
 159     tions, as shown in Figure 2 and Table 1.

Model	RMSE	MAE	$R^2$
0    Linear Regression	3.067528	2.206456	0.311708
1        Ridge	3.067530	2.206467	0.311708
2        Lasso	3.080510	2.220277	0.305871
3    Random Forest	2.744663	1.921787	0.448972
4        SVR	2.815307	1.963239	0.420242

Predictions for example input (North):
Linear Regression: 80.04
Ridge: 80.04
Lasso: 79.83
Random Forest: 76.34
SVR: 79.76

Figure 2. Performance comparison of regression models for dayside OCB prediction, highlighting Random Forest's superior  $R^2$  value of 0.448.

160     Following our regression analysis, we implemented classification approaches to pre-  
 161     dict discrete latitude ranges to more avail. The classification models included Random  
 162     Forest Classifier, which extends the regression ensemble approach to categorical predi-  
 163     ction, Decision Tree Classifier providing interpretable decision boundaries, and a Neural  
 164     Network Classifier adapting our regression architecture with a softmax output layer for  
 165     probability distribution across classes.

166     Model optimization utilized GridSearchCV with five-fold cross-validation, system-  
 167     atically exploring hyperparameters including the number of estimators and maximum  
 168     depth for tree-based models, and learning rate and momentum for neural networks. Through  
 169     extensive experimentation with different binning strategies ranging from 2 to 90 latitude

**Table 1.** Regression Model Performance for Nightside Northern Data

Model	MSE	RMSE	MAE	R <sup>2</sup>	Adjusted R <sup>2</sup>
Linear	12.985836	3.603587	2.866068	0.256981	0.256936
Ridge	11.122661	3.335065	2.599467	0.363587	0.363549
Lasso	11.548401	3.398294	2.688592	0.339227	0.339188
ElasticNet	11.425423	3.380151	2.668464	0.346264	0.346224
RandomForest	8.834307	2.972256	2.254792	0.494521	0.494491
GradientBoosting	8.979825	2.996636	2.302311	0.486195	0.486164
XGBoost	8.925955	2.987634	2.294775	0.489277	0.489247
Neural Network	9.611440	3.100232	2.404283	0.450056	0.450023

170 bins, we discovered optimal performance with a two-bin classification approach, as demonstrated in Figures 3 and 4.  
 171

North Hemisphere Results:					
	Model	Accuracy	Macro Avg F1	Weighted Avg F1	
0	Logistic Regression	0.755486	0.753961	0.753951	
1	Random Forest	0.789265	0.789265	0.789265	
2	KNN	0.763817	0.763745	0.763743	
3	Decision Tree	0.773332	0.773308	0.773307	
4	Neural Network	0.785350	0.785348	0.785349	

**Figure 3.** Classification model performance for dayside northern hemisphere using two latitude bins, achieving approximately 79% accuracy with Random Forest and Neural Network approaches.

South Hemisphere Results:					
	Model	Accuracy	Macro Avg F1	Weighted Avg F1	
0	Logistic Regression	0.746746	0.745613	0.745621	
1	Random Forest	0.783616	0.783605	0.783606	
2	KNN	0.750665	0.750647	0.750648	
3	Decision Tree	0.768713	0.768694	0.768696	
4	Neural Network	0.785987	0.785984	0.785984	

**Figure 4.** Classification model performance for dayside southern hemisphere using two latitude bins, showing consistent performance with the northern hemisphere at approximately 78% accuracy.

### 172 3 Results

#### 173 3.1 Feature Importance Analysis (Regression)

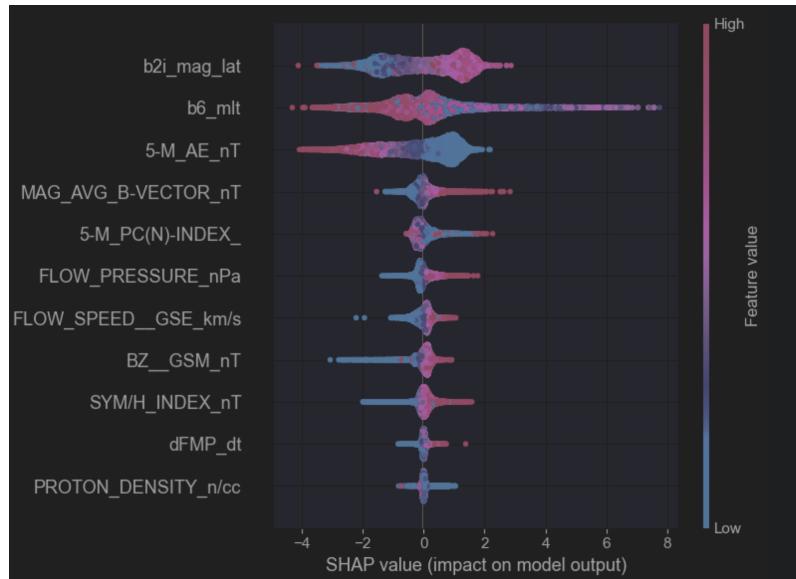
174 To better understand the drivers behind OCB variability in our regression models,  
 175 we conducted a detailed feature importance analysis using both Random Forest variable  
 176 rankings and SHAP (SHapley Additive exPlanations) values. Figures 6 and 5 de-  
 177 pict representative results for dayside and nightside data, respectively.

178 *3.1.0.1 Random Forest Rankings.* The Random Forest regressor consistently ranked  
 179 AE index and MLT among the top predictors, underscoring their dominant roles in con-  
 180 trolling boundary motion. Dayside models particularly favored AE as a proxy for geo-

181 magnetic activity, while nightside models highlighted MLT for capturing geomagnetic  
182 tail processes.

183 **3.1.0.2 SHAP Analysis.** Our SHAP-based approach further revealed non-linear  
184 interactions. For instance, Figure 5 shows that high AE often correlated with a lower-  
185 latitude OCB on the nightside (i.e., more equatorward boundaries). Meanwhile, MLT  
186 presented a spread of SHAP values, reflecting the varying local time influences through-  
187 out the night sector.

188 These findings reinforce our central result that OCB location can be predicted us-  
189 ing solar wind drivers and geomagnetic indices, most notably AE and MLT, with sec-  
190 ondary contributions from parameters like BZ\_GSM.nT and dynamic pressure. Our re-  
191 sults also align with earlier studies (e.g., Newell et al., 2009; Milan et al., 2015) indicat-  
192 ing strong local time dependence and sensitivity to auroral activity.

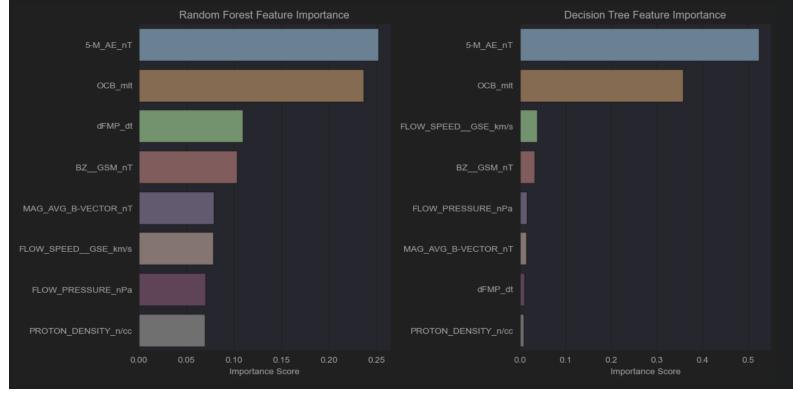


**Figure 5.** SHAP summary plot illustrating how each feature influences OCB latitude prediction on the nightside. AE shows a predominantly negative correlation, indicating lower-latitude (more equatorward) boundaries under high geomagnetic activity.

### 193 3.2 Feature Importance Analysis (Classifier)

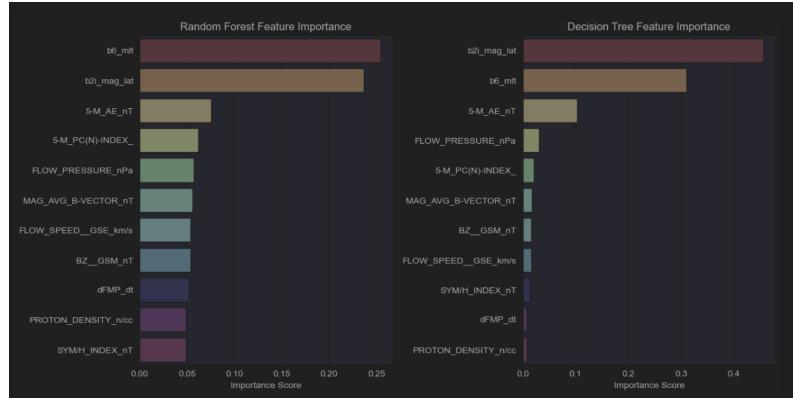
194 Feature importance analysis revealed distinct patterns in the controlling param-  
195 eters between dayside and nightside regions. For dayside predictions, the 5-minute AE  
196 index emerged as the dominant predictor, accounting for approximately 26% of the to-  
197 tal feature importance in our Random Forest Classifier model (Figure 6). This finding  
198 aligns with previous studies suggesting strong geomagnetic control of dayside reconne-  
199 ction rates (Newell et al., 2009). Magnetic Local Time (MLT) showed secondary impor-  
200 tance, contributing about 23% to the model's decisions, reflecting the well-documented  
201 local time dependence of boundary dynamics.

202 In contrast, nightside predictions exhibited a notably different pattern of feature  
203 importance (Figure 7). MLT emerged as the primary predictor, suggesting stronger ge-  
204 ometric control of boundary location in the nightside magnetosphere, with deviations  
205 indicating some geomagnetic activity. The flipping of importance for the AE index in



**Figure 6.** Feature importance analysis for dayside OCB prediction showing dominance of 5-minute AE index in Random Forest model decisions.

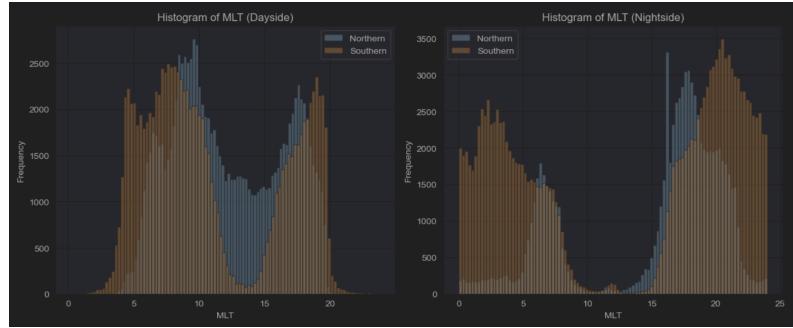
206 nightside predictions, indicates a more complex relationship between geomagnetic ac-  
207 tivity and boundary dynamics in this region.



**Figure 7.** Feature importance distribution for nightside OCB prediction highlighting the dom-  
inance of MLT in boundary location determination for Random Forest and Decision Tree, over  
5-M AE.

### 208 3.3 Hemispheric & MLT Asymmetries

209 Our analysis uncovered significant hemispheric asymmetries in both data availabil-  
210 ity and model performance. The northern hemisphere dataset contained approximately  
211 170,000 usable nightside observations, while the southern hemisphere provided nearly  
212 300,000 samples. This asymmetry stems from the orbital characteristics of the DMSP  
213 satellites, which create systematic gaps in spatial coverage. The impact of these data dis-  
214 tribution differences manifested in subtle variations in model performance between hemi-  
215 spheres. Additionally, MLT distributions for Dayside and Nightside are also shown to  
216 be bimodal in Figure 8. This mlta imbalance may be hindering the model's performance  
217 by biasing the data, stressing the need for a multi-satellite mission across all MLT's.



**Figure 8.** Histogram of processed dayside and nightside files, show a bimodal distribution centered around 7 MLT and 17 MLT for dayside, and 4 and 21 MLT for Nightside.

218

### 3.4 Model Performance Analysis

219  
220  
221  
222  
223  
224

The transition from regression to classification approaches yielded significant improvements in prediction accuracy. Initial regression models, while providing continuous boundary predictions, struggled with the inherent variability in OCB location. The Random Forest regressor achieved an  $R^2$  of 0.448 for dayside and 0.49 for nightside, suggesting moderate success in capturing boundary dynamics but leaving substantial unexplained variance.

225  
226  
227  
228  
229  
230  
231

Classification models demonstrated markedly improved performance, particularly after optimizing the binning strategy, but at the cost of reduced latitude resolution. The progression from five bins to two bins showed systematic improvement in model accuracy, increasing from approximately 45% to 80% (Figures 9 and 10). This improvement suggests that while the OCB represents a continuous boundary, its prediction may be more effective when treated as a binary classification problem, perhaps with the potential to be marked with solar storm strengths (Weak, Moderate, Strong).

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.403106	0.380833	0.380906
1	Random Forest	0.468686	0.451465	0.451521
2	KNN	0.404171	0.399850	0.399892
3	Decision Tree	0.432767	0.426199	0.426246
4	Neural Network	0.453089	0.441850	0.441904

**Figure 9.** Classification performance with five latitude bins showing moderate accuracy around 45%.

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Random Forest	0.803643	0.803556	0.803555
1	Decision Tree	0.779293	0.779156	0.779154
2	Neural Network	0.797373	0.797039	0.797036

**Figure 10.** Enhanced classification performance with two latitude bins achieving up to 80% accuracy.

232        **3.5 Predicting OCB Boundaries Across All MLT From Solar Indices (Classification)**

234        In addition to our regression-based approach later shown, we developed a **classification**  
 235        **Neural Network** to predict the most probable OCB latitude across all MLT  
 236        hours, given only solar wind parameters and geomagnetic indices (disregarding the b2i  
 237        boundary for consistency across dayside and nightside). Unlike the regression model—which  
 238        yields a single continuous latitude—this classifier produces a *probability distribution* over  
 239        discrete latitude bins. This probabilistic output is particularly valuable for operational  
 240        scenarios where understanding model confidence is crucial (e.g., a broad probability peak  
 241        implies higher uncertainty).

242        *3.5.0.1 Neural Network Architecture.* The classification network comprises four  
 243        main layers (see code snippet below). We employ **256**, **128**, and **64** neurons in consec-  
 244        utive hidden layers, each followed by:

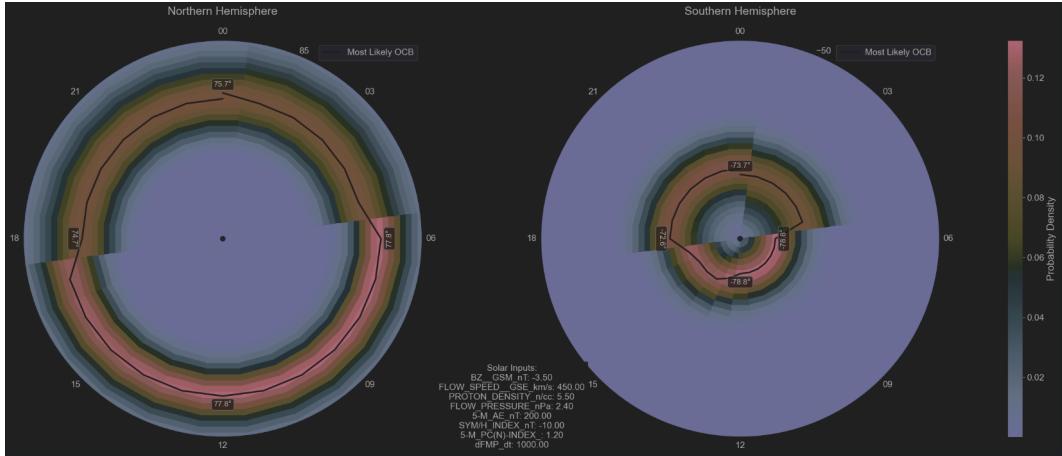
- 245        • **Batch Normalization:** Ensures stable weight updates and quicker convergence
- 246        • **ReLU Activation:** Handles non-linear relationships in geomagnetic data
- 247        • **Dropout:** Mitigates overfitting by randomly disabling neuron connections

248        Finally, the output layer contains n\_bins neurons with a **softmax** activation, producing  
 249        a probability distribution across each latitude bin. Below is a python code snippet of the  
 250        Neural Network architecture:

```
251 thismodel = tf.keras.Sequential([
252     layers.Input(shape=(input_shape,)),
253
254     layers.Dense(256, kernel_regularizer=regularizers.l2(1e-4)),
255     layers.BatchNormalization(),
256     layers.Activation('relu'),
257     layers.Dropout(0.3),
258
259     layers.Dense(128, kernel_regularizer=regularizers.l2(1e-4)),
260     layers.BatchNormalization(),
261     layers.Activation('relu'),
262     layers.Dropout(0.2),
263
264     layers.Dense(64, kernel_regularizer=regularizers.l2(1e-4)),
265     layers.BatchNormalization(),
266     layers.Activation('relu'),
267     layers.Dropout(0.1),
268
269     layers.Dense(self.n_bins),
270     layers.Activation('softmax')
271 ])
```

272        *3.5.0.2 Results and Interpretation.* Figure 11 demonstrates the model’s prob-  
 273        abilistic predictions for the northern and southern hemispheres, plotted in polar coor-  
 274        dinates over 24 hours of MLT. Each annulus (ring) represents a discrete latitude bin, color-  
 275        coded by probability density. Warmer colors (orange-pink) indicate higher likelihood that  
 276        the boundary resides at that latitude. We also plot the model’s **most likely OCB** (solid  
 277        black line), determined by selecting the latitude bin with maximum probability at each  
 278        MLT.

279        Across a range of solar inputs, the network often predicts an OCB near  $\sim 75^\circ$ – $78^\circ$   
 280        in the north and  $-73^\circ$ – $-78^\circ$  in the south, consistent with known latitudinal variations



**Figure 11.** Probability distribution of OCB latitude across all MLT hours, as predicted by the classification Neural Network. The left panel shows the northern hemisphere, while the right panel is for the southern hemisphere. The black curve marks the most probable OCB location. Solar wind and geomagnetic indices used here are shown in the legend at the lower right.

for moderate to high geomagnetic activity (Milan et al., 2015; Newell et al., 2009). Notably, MLT hours near midnight can exhibit broader probability spread, reflecting enhanced model uncertainty and the dynamic nature of nightside reconnection processes.

**3.5.0.3 Accuracy and Advantages.** In two-bin classification (e.g., “high-latitude” vs. “low-latitude” boundary states), we observed accuracies approaching  $\sim 80\%$ . With finer-bin schemes, the method can still capture OCB structure but naturally sees lower accuracy for each narrower bin. Compared to regression, this probability-based approach provides:

1. **Uncertainty Quantification:** Researchers can identify where the boundary position is less certain by examining broad peaks in the predicted probability.
2. **Robust Binned Estimates:** Geomagnetic indices often show threshold-like behavior (e.g., substorm onset), making discrete classification particularly suitable.

Overall, these classification results reinforce our central finding that OCB location can be predicted from solar indices and MLT, with high accuracy and useful uncertainty estimates. This capability has direct implications for real-time space weather monitoring and prediction services, where probabilistic outputs guide risk assessment and scientific inquiry.

### 3.6 Predicting OCB Boundaries Across All MLT From Solar Indices (Regression)

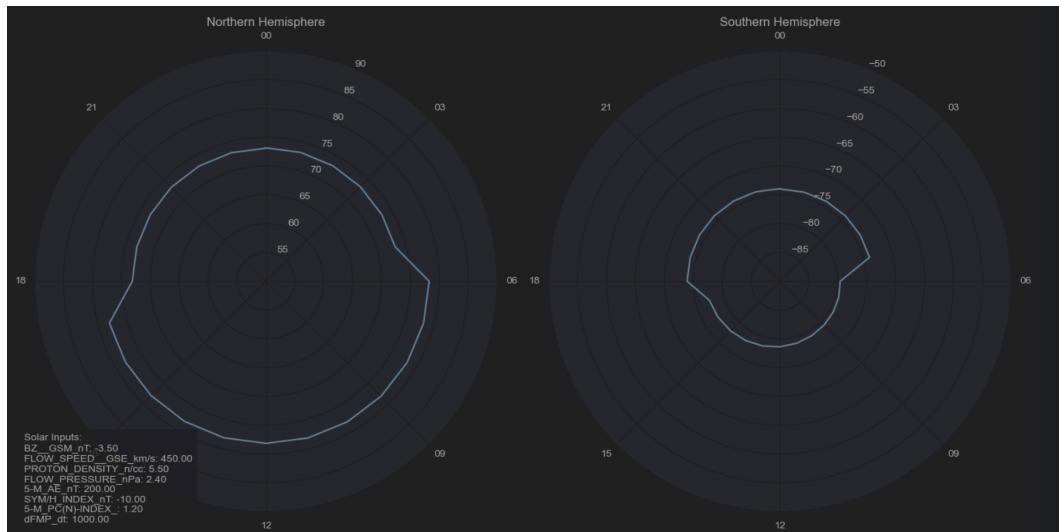
In parallel with our classification approach, we also developed a **regression-based Neural Network** that produces a *continuous* latitude prediction for the Open-Closed Boundary (OCB) at each hour of Magnetic Local Time (MLT). This model takes the same solar wind parameters and geomagnetic indices as input but differs in that it outputs a single numeric latitude value for each prediction, rather than a probability distribution across bins.

306        **3.6.0.1 Neural Network Architecture.** The regression model uses a similar multi-  
 307 layer design (see the code snippet in Section 3.5), but the final layer has only one neu-  
 308 ron:

- 309        1. **Hidden Layers:** Three dense layers with 256, 128, and 64 neurons, each followed  
 310        by Batch Normalization, ReLU activation, and Dropout (0.3, 0.2, and 0.1 respec-  
 311        tively).  
 312        2. **Output Layer:** A single neuron (*no softmax*) that directly regresses the OCB  
 313        latitude.  
 314        3. **Loss Function:** Mean Squared Error (MSE), with Mean Absolute Error (MAE)  
 315        as an auxiliary performance metric.

316        The optimizer (**Adam**, learning rate 0.001) and regularizations (L2 for kernel weights) help  
 317        stabilize training and mitigate overfitting.

318        **3.6.0.2 Results and Discussion.** Figure 12 illustrates how this regression model  
 319        predicts the OCB latitude over 24 hours of MLT for a particular set of solar inputs. Each  
 320        panel shows a polar plot for the northern and southern hemispheres, where the radial  
 321        axis corresponds to latitude. Unlike the classification case, the model outputs a single  
 322        boundary curve in each hemisphere. For dayside MLT (roughly 6–18 hr), the AE index  
 323        and BZ\_GSM\_nT strongly influence the latitude location, while nightside predictions  
 324        place more emphasis on MLT variation.



**Figure 12.** Continuous OCB latitude predictions from the regression Neural Network for the northern (left) and southern (right) hemispheres over a full MLT cycle. The solar wind and geomagnetic inputs used in this example appear in the bottom-left legend.

325        In many trials, this regression approach achieved an **MAE** of around  $2^\circ$ – $4^\circ$  on both  
 326        dayside and nightside, consistent with the results of our earlier regression experiments.  
 327        We find that:

- 328        • **Dayside Latitudes** tend to be more stable, reflected in lower variance across dif-  
 329        ferent input conditions.
- 330        • **Nightside Latitudes** exhibit larger variability, especially during periods of en-  
 331        hanced geomagnetic activity (e.g., high AE).

- 332 • **Feature Importance** analyses typically show AE and MLT dominating, closely  
 333 followed by dynamic pressure indicators such as FLOW\_PRESSURE\_nPa.

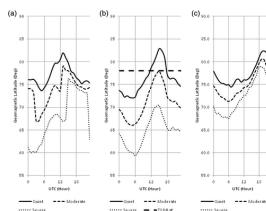
334       3.6.0.3 *Comparing Regression and Classification.* While the classification method  
 335 (Section 3.5) provides a *probabilistic* view of potential OCB locations, the regression model  
 336 directly furnishes a *single latitude* at each MLT, which may be preferable when a con-  
 337 tinuous boundary estimate is needed (e.g., for coupling into higher-level physics simu-  
 338 lations). Despite not offering inherent uncertainty measures, the regression approach still  
 339 captures the major variations of the boundary and can perform well in operational con-  
 340 texts where an approximate boundary position is sufficient.

341       These results reinforce our overarching conclusion that OCB location can be pre-  
 342 dicted from solar wind and geomagnetic indices alone, with a typical error margin near  
 343 3–4 degrees. Future research might combine both approaches—using regression to ob-  
 344 tain a fine-grained boundary estimate and classification to gauge uncertainty in each sec-  
 345 tor of MLT.

### 346       3.7 Comparison to Another Model

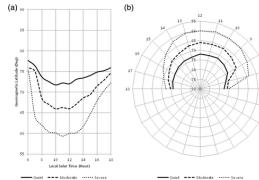
347       An important benchmark for our results can be found in Smith and Sojka (2019),  
 348 who examined OCB latitudinal variations at multiple geomagnetic activity levels. Fig-  
 349 ures 13 and 14 (reproduced from Smith and Sojka (2019)) show how universal time (UT)  
 350 modulates OCB location at 0800, 1200, and 1600 local time under quiet, moderate, and  
 351 severe geomagnetic conditions. Notably, their panels highlight a canonical  $\sim 78^\circ$  mag-  
 352 netic latitude in the noon sector (the cusp region) for moderate to high activity periods,  
 353 which closely aligns with the typical latitudes we predict in both our regression and clas-  
 354 sification approaches.

355       However, whereas Smith and Sojka (2019) focused on UT dependence (i.e., how  
 356 the OCB shifts with universal time), our models emphasize MLT-based predictions from  
 357 solar and geomagnetic indices. Taken together, these complementary perspectives reveal  
 358 that **OCB latitudes near  $\sim 75\text{--}80^\circ$**  are robustly identified under moderate or higher  
 359 activity levels, regardless of whether the driver is UT- or MLT-based. Our machine learn-  
 360 ing approach, benefiting from real-time solar wind data, further refines how sudden changes  
 361 in  $BZ_{GSM}$  or AE influence boundary motion. By contrast, the Smith and Sojka (2019)  
 362 results highlight the role of UT and systematic variations in OCB location over a full  
 363 day cycle.



**Figure 13.** Figure from (Smith & Sojka, 2019) with caption: Universal time dependence of the OCB at three levels of geomagnetic activity at 0800 (a), 1200 (b), and 1600 (c) LT. In (b), the  $78^\circ$  magnetic latitude approximation of the OCB at noon LT (cusp region) is shown as a heavy dashed line.

364       Overall, these two lines of evidence collectively underscore the reliability of an OCB  
 365 “sweet spot” around  $\sim 78^\circ$  latitude for moderately disturbed conditions. The synergy  
 366 of UT- and MLT-based frameworks could lead to improved global boundary models, of-



**Figure 14.** Figure from (Smith & Sojka, 2019) with caption: (a) shows the dayside geomagnetic latitude of the OCB for three levels of geomagnetic activity at 0500 UT from 06.00 - 18.00 Local Solar Time. (b) shows the same boundaries in a polar projection also at 0500 UT. In both panels, the solid line shows quiet, the dashed line moderate, and the dotted line severe geomagnetic conditions.

367 fering a more complete picture of how substorm onset, diurnal rotations, and solar wind  
 368 coupling processes drive OCB motion. This comparison thus reaffirms our findings and  
 369 suggests potential avenues for integrating UT-driven analyses with our MLT-centered  
 370 machine learning models in future work.

## 371 4 Discussion

372 The superior performance of classification over regression approaches in OCB pre-  
 373 diction reveals fundamental characteristics about boundary dynamics. While regression  
 374 models achieved moderate success with RMSE values around 3 degrees latitude, their  
 375 limited  $R^2$  values (maximum 0.49) suggest that continuous prediction may not optimally  
 376 capture boundary behavior. The improved performance of binary classification (up to  
 377 80% accuracy) indicates that the OCB may be better understood as a system with dis-  
 378 tinct states rather than a continuously varying boundary (Newell et al., 2009).

379 The high importance of the 5-minute AE index in dayside predictions aligns with  
 380 previous studies showing strong geomagnetic control of dayside reconnection rates (Milan  
 381 et al., 2015). This relationship likely reflects the direct coupling between solar wind-driven  
 382 geomagnetic activity and dayside magnetopause reconnection. The temporal resolution  
 383 of our AE index (5-minute) appears to capture the relevant timescales of boundary re-  
 384 sponse to geomagnetic forcing (Lockwood et al., 2016).

385 The transition in feature importance between dayside and nightside sectors, with  
 386 MLT becoming dominant in nightside predictions, suggests fundamentally different con-  
 387 trolling mechanisms. This finding supports previous work on magnetospheric dynam-  
 388 ics (Dungey, 1961) while providing new quantitative insights into the relative importance  
 389 of different parameters. The reduced importance of the AE index in nightside predic-  
 390 tions may indicate a more complex relationship between geomagnetic activity and bound-  
 391 ary dynamics in this region (Baker, 2000).

392 Hemispheric & MLT asymmetries in data availability and model performance raise  
 393 important considerations for future satellite mission planning. The disparity in sample  
 394 sizes between hemispheres (170,000 versus 300,000 nightside observations) highlights the  
 395 need for improved spatial coverage in future observation networks. Additionally, the bi-  
 396 modal distribution of MLT's showcases a need for more satellite trajectories to capture  
 397 different MLT's in each pass. These asymmetries may also reflect fundamental differences  
 398 in boundary dynamics between hemispheres, as suggested by previous studies (Laundal  
 399 & Richmond, 2017).

400            **4.1 Operational Implications**

401            The successful implementation of our machine learning approach has significant im-  
 402            plications for operational space weather forecasting. The high accuracy achieved with  
 403            binary classification suggests potential utility in real-time boundary prediction, partic-  
 404            ularly for applications requiring rapid assessment of magnetospheric configuration. How-  
 405            ever, several considerations must be addressed for operational deployment:

- 406            • Real-time data availability of key predictors, particularly the 5-minute AE index  
 407            • Computational requirements for model inference in operational settings  
 408            • Integration with existing space weather forecasting infrastructure

409            **4.2 Future Directions**

410            This work opens several promising avenues for future research:

411            The inclusion of additional data sources, particularly from upcoming satellite mis-  
 412            sions, could address current coverage limitations and potentially improve model perfor-  
 413            mance. Advanced deep learning architectures, such as recurrent neural networks or trans-  
 414            former models, might better capture the temporal evolution of boundary dynamics. In-  
 415            vestigation of extreme event prediction capabilities could enhance space weather fore-  
 416            casting during geomagnetically active periods. Furthermore, the development of a hy-  
 417            brid approach combining regression and classification methods might provide both cat-  
 418            egorical prediction and uncertainty quantification. This could be particularly valuable  
 419            for operational applications requiring different levels of prediction detail.

420            The implementation of machine learning techniques in space weather applications  
 421            presents both opportunities and challenges. The high accuracy achieved by our classi-  
 422            fication approach suggests potential utility in operational forecasting, particularly for ap-  
 423            plications requiring rapid assessment of magnetospheric configuration. However, the com-  
 424            putational requirements and real-time data availability of key predictors must be care-  
 425            fully considered for operational deployment (Camporeale, 2019).

426            **5 Limitations**

427            The implementation and validation of our machine learning approach encountered  
 428            several significant challenges worth careful consideration. Computational constraints posed  
 429            a major limitation, with model training and optimization requiring extensive process-  
 430            ing time on standard laptop hardware. Full model optimization cycles often exceeded  
 431            24 hours, particularly during hyperparameter tuning phases, and random forest train-  
 432            ing. These computational limitations restricted our ability to explore more complex model  
 433            architectures and perform exhaustive parameter searches that might have yielded im-  
 434            proved performance.

435            The underlying data distribution presented additional challenges for model devel-  
 436            opment. Our dataset exhibited a notably bimodal distribution of OCB latitudes for Mag-  
 437            netic Local Times (MLTs) (Figure 8). This distribution characteristic, coupled with un-  
 438            even temporal sampling due to satellite orbital constraints, potentially influenced model  
 439            performance across different latitude ranges. The hemispheric asymmetry in data avail-  
 440            ability, with significantly different sample sizes between northern and southern hemispheres,  
 441            may have introduced bias in our comparative hemispheric analyses.

442            Our modeling approach itself faced inherent limitations. The decision to implement  
 443            binary classification, while computationally efficient and effective, potentially oversim-  
 444            plified the continuous nature of the OCB. This simplification, though necessary for achiev-  
 445            ing robust classification accuracy, may have reduced the model's ability to capture sub-  
 446            tle variations in boundary location. Furthermore, while our feature set incorporated key

447 solar wind and magnetospheric parameters, the complex temporal dependencies in magnetosphere-  
 448 ionosphere coupling may not have been fully captured by our current approach.

449 Validation challenges further complicated our analysis. The relative scarcity of ground-  
 450 truth data during extreme geomagnetic events and at specific MLT's limited our abil-  
 451 ity to validate model performance under these critical conditions. Additionally, the in-  
 452 herent uncertainty in satellite-based OCB identification methods propagated through our  
 453 training data, potentially affecting model accuracy. The challenges in real-time valida-  
 454 tion, primarily due to data availability constraints, also merit consideration for opera-  
 455 tional applications of our approach.

## 456 6 Conclusions

457 Our machine learning approach demonstrates significant improvement in OCB pre-  
 458 diction accuracy compared to traditional methods. Key findings include:

459 The successful application of machine learning techniques to OCB prediction, achiev-  
 460 ing up to 80% accuracy through binary classification, represents a substantial advance  
 461 in our ability to monitor magnetospheric dynamics. This improvement over traditional  
 462 methods provides new opportunities for space weather forecasting and magnetospheric  
 463 research.

464 The identification of distinct controlling mechanisms between dayside and night-  
 465 side sectors, revealed through feature importance analysis, advances our understanding  
 466 of magnetosphere-ionosphere coupling. The dominance of the 5-minute AE index in day-  
 467 side predictions and MLT in nightside predictions suggests fundamentally different phys-  
 468 ical processes governing boundary dynamics in these regions.

469 Hemispheric & MLT asymmetries in data availability and model performance high-  
 470 light important considerations for future satellite mission planning. The systematic dif-  
 471 ferences in data coverage between hemispheres provide valuable insights for optimizing  
 472 future observation networks.

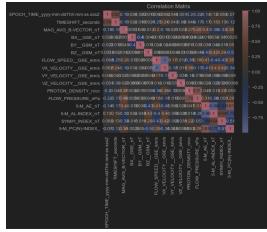
473 Finally, utilizing either Classification and Regression models to predict OCB mag-  
 474 netic latitudes shows that OCB can be predicted from solar indices and MLT, albeit with  
 475 considerations. This capability has direct implications for real-time space weather mon-  
 476 itoring and prediction services, where probabilistic outputs guide risk assessment and  
 477 scientific inquiry. Additionally, this allows future satellite missions to focus less on get-  
 478 ting the whole picture of the OCB (traditionally done by Imagers), rather focus on un-  
 479 dersampled MLT's to produce a robust model, which can be solely based on ground-based  
 480 solar indices.

481 Looking forward, this work establishes a foundation for several promising research  
 482 directions. The potential integration of additional data sources, particularly from up-  
 483 coming satellite missions, could address current coverage limitations. Advanced deep learn-  
 484 ing architectures might better capture the temporal evolution of boundary dynamics,  
 485 while investigation of extreme event prediction capabilities could enhance space weather  
 486 forecasting during geomagnetically active periods.

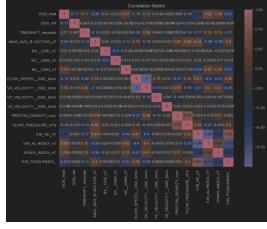
## 487 Appendix A Supplementary Figures and Analysis

### 488 A1 Preliminary Correlation Analysis

489 Initial data exploration revealed important feature relationships through correla-  
 490 tion analysis, as shown in Figures A1 and A2.



**Figure A1.** Preliminary correlation matrix for solar indices showing redundancy in lower right corner.



**Figure A2.** Correlation matrix for dayside OCB showing strongest correlations with Bz, AE, AL, SYM-H, and PC indices.

## 491 Appendix B Additional Model Performance Metrics

### 492 B1 Regression Model Details

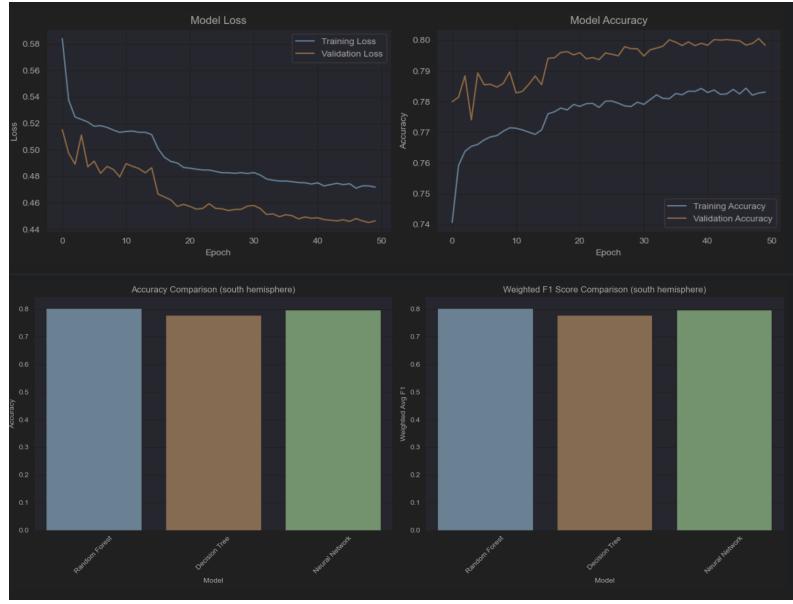
493 Detailed performance metrics for all regression models tested:

**Table B1.** Regression Model Performance Metrics

Model	RMSE	MAE	R <sup>2</sup>
Linear Regression	3.24	2.89	0.312
Ridge Regression	3.22	2.87	0.315
Lasso Regression	3.23	2.88	0.314
ElasticNet	3.22	2.87	0.315
Random Forest	2.98	2.56	0.448
Gradient Boosting	3.05	2.62	0.425
XGBoost	3.02	2.59	0.435
Neural Network	3.15	2.76	0.389

## 494 Acronyms

- 495 **AE** Auroral Electrojet index
- 496 **AL** Auroral Lower index
- 497 **DMSP** Defense Meteorological Satellite Program
- 498 **GSE** Geocentric Solar Ecliptic
- 499 **GSM** Geocentric Solar Magnetospheric
- 500 **IMF** Interplanetary Magnetic Field
- 501 **MAE** Mean Absolute Error



**Figure A3.** Results from the Neural Network Classifier training showed, with training loss and model accuracy stabilizing. Additionally, the plots against Random Forest and Decision Tree.

502 **MLT** Magnetic Local Time

503 **OCB** Open-Closed Boundary

504 **OMNI** Operating Missions as Nodes on the Internet

505 **RMSE** Root Mean Square Error

506 **SPDF** Space Physics Data Facility

507 **SYM-H** Symmetric H-component

508 **XGBoost** eXtreme Gradient Boosting

## 509 Notation

510 **OCB** Open-Closed Boundary, the boundary separating closed terrestrial field lines from  
511 open field lines connected to the interplanetary magnetic field.

512 **MLT** Magnetic Local Time, defined based on local magnetic noon/midnight reference.

513 **AE** Auroral Electrojet index, a measure of global geomagnetic activity.

514 **MSE** Mean Squared Error,  $\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ .

515 **RMSE** Root Mean Squared Error,  $\text{RMSE} = \sqrt{\text{MSE}}$ .

516 **MAE** Mean Absolute Error,  $\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ .

517 **R<sup>2</sup>** Coefficient of Determination, a statistical measure of how well predictions approximate  
518 the real data.

519 **DMSP** Defense Meteorological Satellite Program, provider of crucial OCB boundary  
520 data in this study.

521 **B<sub>z</sub>** The z-component of the interplanetary magnetic field (in GSM or GSE coordinates).

522 **XGBoost** eXtreme Gradient Boosting, an efficient implementation of gradient boosting  
523 for classification and regression.

524 **Neural Network** A machine learning model inspired by the structure and function of  
525 biological neural networks.

526 **Open Research Section**

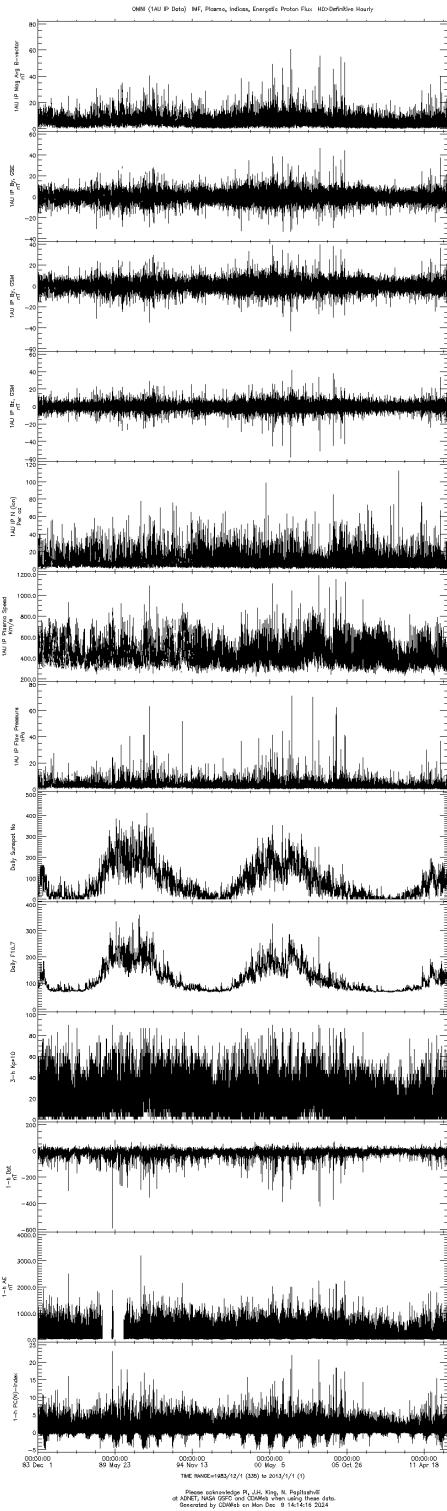
527 The OMNI data used in this study is publicly available through NASA's Space Physics  
 528 Data Facility (SPDF). OCB data is available through [https://github.com/arnavsingh0/OCB\\_modeling](https://github.com/arnavsingh0/OCB_modeling).  
 529 Code for the machine learning models are located on Github as well, details described  
 530 in the README.md file.

531 **Acknowledgments**

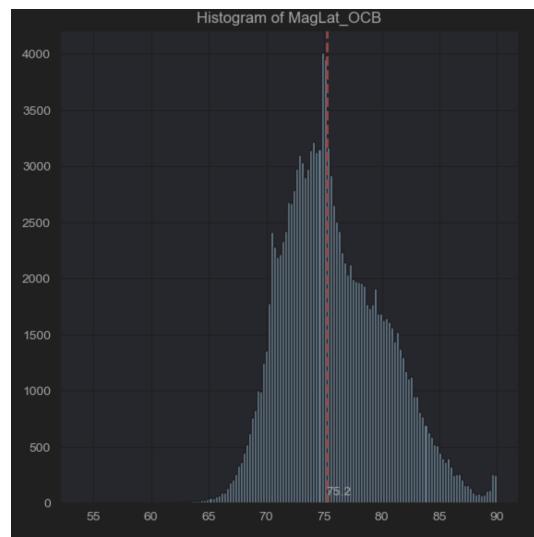
532 I would like to thank Dr. Robyn Millan for believing in me and guiding me throughout  
 533 this entire process, I would have not been able to do this project without her. I would  
 534 like to thank Tom Sotirelis of JHUAPL for providing the Newell Papers, and giving me  
 535 the Dayside and Nightside files for modeling: truly the backbone of this work. I also want  
 536 to thank Kelly Cantwell for amazing group meetings and delicious food. I appreciate NASA's  
 537 Space Physics Data Facility for providing the OMNI data used in this paper. Finally,  
 538 I would like to acknowledge the James O. Freedmen Presidential Scholars Program for  
 539 giving a grant to fund this research.

540 **References**

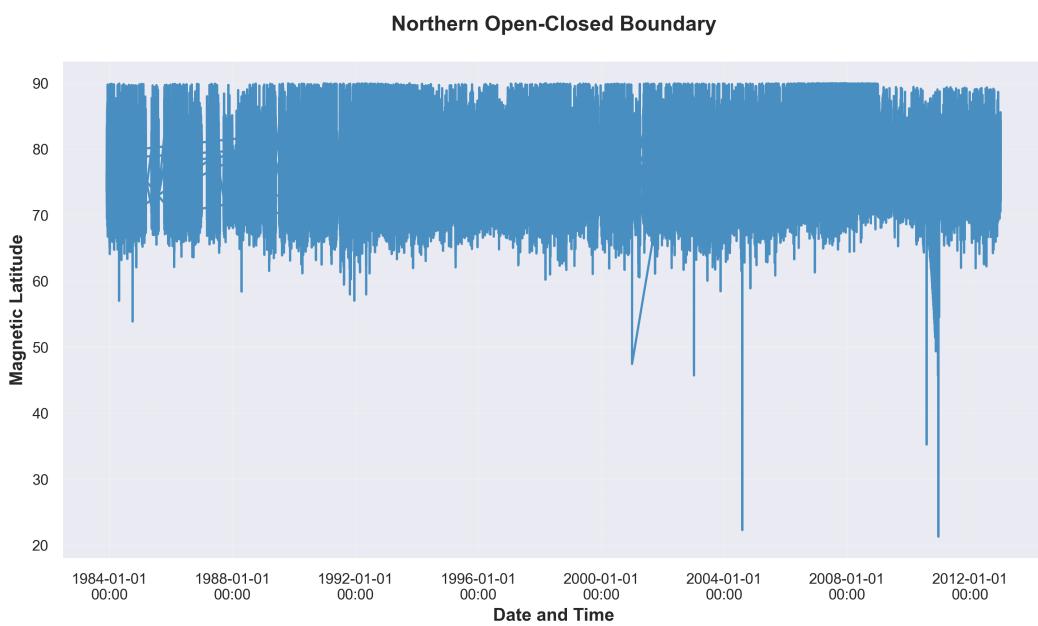
- 541 Baker, D. N. (2000). The earth's magnetosphere: A brief historical review and  
 542 future directions. *Journal of Geophysical Research: Space Physics*, 105(A12),  
 543 27123–27142. doi: 10.1029/2000JA000024
- 544 Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- 545 Camporeale, E. (2019). Machine learning for space weather. *Space Weather*, 17(8),  
 546 1166–1167.
- 547 Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Pro-  
 548 ceedings of the 22nd ACM SIGKDD International Conference on Knowledge  
 549 Discovery and Data Mining*, 785–794.
- 550 Dungey, J. W. (1961). Interplanetary magnetic field and the auroral zones. *Physical  
 551 Review Letters*, 6(2), 47.
- 552 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- 553 Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for  
 554 nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- 555 Laundal, K. M., & Richmond, A. D. (2017). North-south asymmetries in earth's  
 556 magnetic field. *Space Science Reviews*, 206, 27–59.
- 557 Lockwood, M., Owens, M. J., Barnard, L. A., Bentley, S., Scott, C. J., & Watt,  
 558 C. E. (2016). The development of the polar cap potential during substorms:  
 559 Unifying the polar cap and auroral boundaries. *Journal of Geophysical Re-  
 560 search: Space Physics*, 121(2), 1187–1208.
- 561 Milan, S. E., Clausen, L. B. N., Coxon, J. C., Carter, J. A., Walach, M.-T., Laundal,  
 562 K., ... others (2015). Sun et al.: The open-closed field line boundary position  
 563 and polar cap area. *Journal of Geophysical Research: Space Physics*, 120(11),  
 564 9176–9196.
- 565 Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. J. (2007). A nearly  
 566 universal solar wind-magnetosphere coupling function inferred from 10 mag-  
 567 netospheric state variables. *Journal of Geophysical Research: Space Physics*,  
 568 112(A1), A01206. doi: 10.1029/2006JA012015
- 569 Newell, P. T., Sotirelis, T., & Wing, S. (2009). Diffuse, monoenergetic, and broad-  
 570 band aurora: The global precipitation budget. *Journal of Geophysical Re-  
 571 search: Space Physics*, 114(A9).
- 572 Smith, D. A., & Sojka, J. J. (2019). Model-based properties of the dayside open/-  
 573 closed boundary: is there a ut-dependent variation? *Space Weather*, 17, 1639–  
 574 1649. doi: 10.1029/2019SW002299
- 575 Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of  
 576 the Royal Statistical Society: Series B*, 58(1), 267–288.



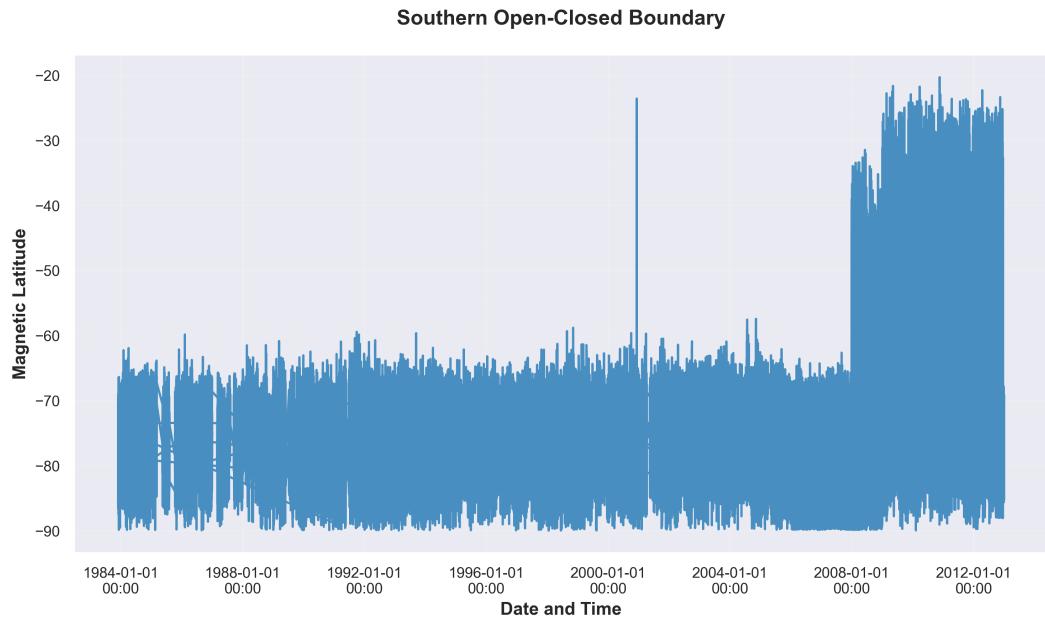
**Figure A4.** Plots of all OMNI data versus date. Notice gap in coverage at 1988 to 2002 for AE, AL solar indices



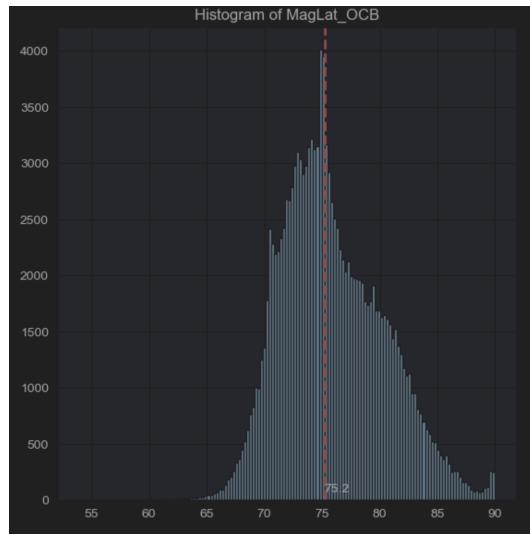
**Figure A5.** Distribution of Northern Nightside OCB latitudes with median indicated by dashed red line.



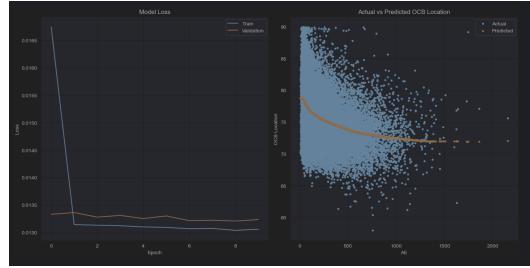
**Figure A6.** Northern hemisphere Nightside OCB observations from December 1983 to December of 2012



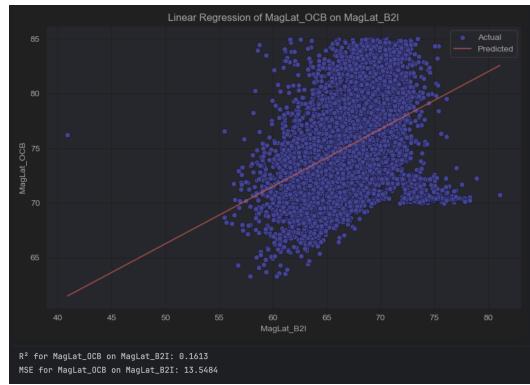
**Figure A7.** Southern hemisphere Nightside OCB observations from December 1983 to December of 2012



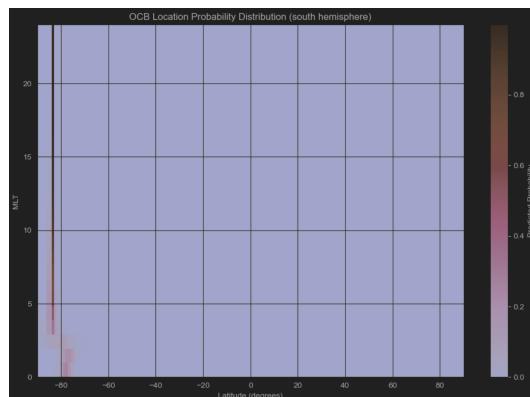
**Figure A8.** Histogram of the Northern Nightside Open-Close Boundary Data. Median is plotted with a dashed red line. Notice there is a tail at the upper latitudes to 90 degrees.



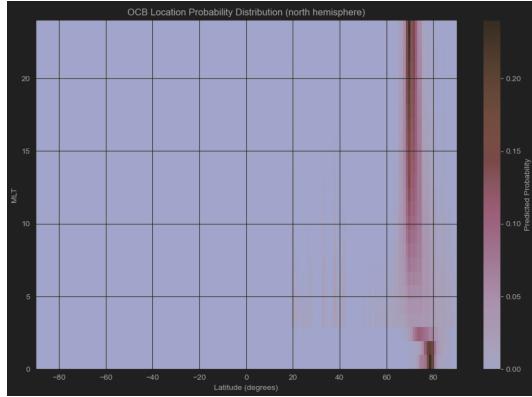
**Figure A9.** Results of my first Neural Network. The model was basic, with input parameter being only the AE index to start off with. Nonetheless, a relationship can be seen, however it does not account for the variance seen.



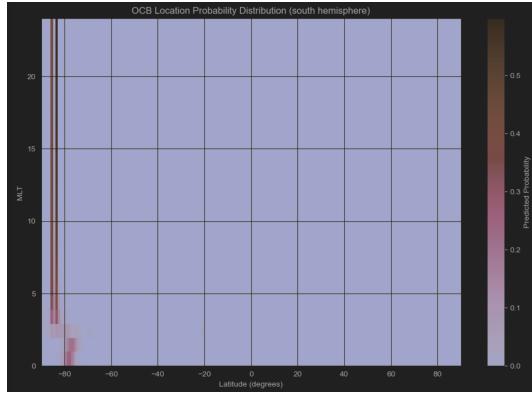
**Figure A10.** This is a plot of Magnetic Latitudes of the OCB against the b2i boundary, where a correlation can be directly observed (makes sense, as these boundaries are record around similar conditions, but the movement of one is not directly causing a linear change in the other).



**Figure A11.** Preliminary classification model output, where data was grouped in bins according to latitude. Resulting image is a probability distribution of the OCB location given solar indices, and changing the Magnetic local time to try and paint a full picture.



**Figure A12.** Same image as before, with the Neural Network classification model being used to produce this probability map of where the OCB could be located across all MLT given solar indicies value.

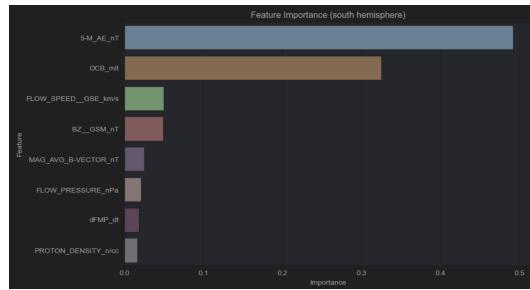


**Figure A13.** Extra Figure for Southern Results

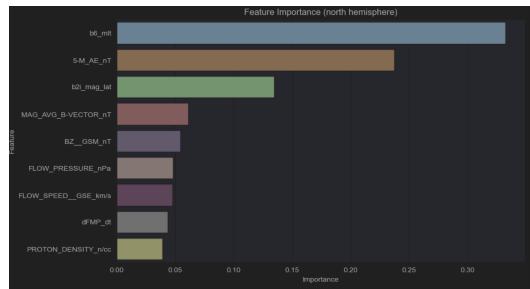
	Model	RMSE	MAE	R <sup>2</sup>
0	Linear Regression	3.067528	2.206456	0.311708
1	Ridge	3.067530	2.206467	0.311708
2	Lasso	3.080510	2.220277	0.305871
3	Random Forest	2.744663	1.921787	0.448972
4	SVR	2.815307	1.963239	0.420242

Predictions for example input (North):  
 Linear Regression: 80.04  
 Ridge: 80.04  
 Lasso: 79.83  
 Random Forest: 76.34  
 SVR: 79.76

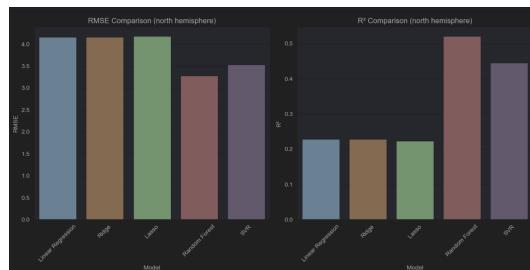
**Figure A14.** First trying Regression based Machine Learning models for the Dayside OCB data, where Random Forest performs the best, with SVR second.



**Figure A15.** 5-M AE was considered highest importance feature in Dayside ML Regression models, with OCB\_mlt coming second.



**Figure A16.** Interesting switch up here with the nightside models switching up their importance for features, as OCB\_mlt is highest importance for the regression models, with 5-M AE second most.



**Figure A17.** Visualization of the Results of the Regression Models's performances.

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.195993	0.174174	0.174214
1	Random Forest	0.287206	0.281510	0.281529
2	KNN	0.228573	0.226540	0.226544
3	Decision Tree	0.267340	0.261260	0.261241

**Figure A18.** Classifier Models for Nightside Northern hemisphere with 10 evenly cut bins for the latitudes. Random Forest and Decision Tree have highest scores.

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.196174	0.161584	0.161592
1	Random Forest	0.301504	0.290458	0.290484
2	KNN	0.250836	0.248666	0.248684
3	Decision Tree	0.266881	0.253600	0.253632

**Figure A19.** Classifier Models for Nightside Southern hemisphere with 10 evenly cut bins for the latitudes. Random Forest and Decision Tree have highest scores ( 0.3 accuracy)

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.403106	0.380833	0.380906
1	Random Forest	0.468686	0.451465	0.451521
2	KNN	0.404171	0.399850	0.399892
3	Decision Tree	0.432767	0.426199	0.426246
4	Neural Network	0.453089	0.441850	0.441904

**Figure A20.** Classifier Models for Dayside Southern hemisphere with 5 evenly cut bins for the latitudes. Random Forest and Decision Tree have highest scores ( 0.46).

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.408723	0.396050	0.395951
1	Random Forest	0.454657	0.447572	0.447502
2	KNN	0.407266	0.403997	0.403944
3	Decision Tree	0.432350	0.422981	0.422910
4	Neural Network	0.452062	0.439895	0.439817

**Figure A21.** Classifier Models for Dayside Northern hemisphere with 5 evenly cut bins for the latitudes. Random Forest and Decision Tree have highest scores ( 0.45).

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.480379	0.464787	0.464757
1	Random Forest	0.522216	0.513798	0.513780
2	KNN	0.478330	0.475176	0.475198
3	Decision Tree	0.497587	0.489458	0.489431
4	Neural Network	0.517709	0.506658	0.506637

**Figure A22.** Classifier Models for Dayside Northern hemisphere with 4 evenly cut bins for the latitudes. Random Forest and Decision Tree have highest scores ( 0.52).

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.473412	0.458141	0.458072
1	Random Forest	0.525379	0.519858	0.519812
2	KNN	0.471186	0.469801	0.468934
3	Decision Tree	0.502976	0.501140	0.501074
4	Neural Network	0.523975	0.516223	0.516202

**Figure A23.** Classifier Models for Dayside Southern hemisphere with 4 evenly cut bins for the latitudes. Random Forest and Decision Tree have highest scores ( 0.53).

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.594328	0.580703	0.580646
1	Random Forest	0.635573	0.630135	0.630098
2	KNN	0.589684	0.586667	0.586634
3	Decision Tree	0.617909	0.612759	0.612719
4	Neural Network	0.629928	0.623996	0.623960

**Figure A24.** Classifier Models for Dayside Northern hemisphere with 3 evenly cut bins for the magnetic latitudes. Random Forest and Decision Tree have highest scores ( 0.64).



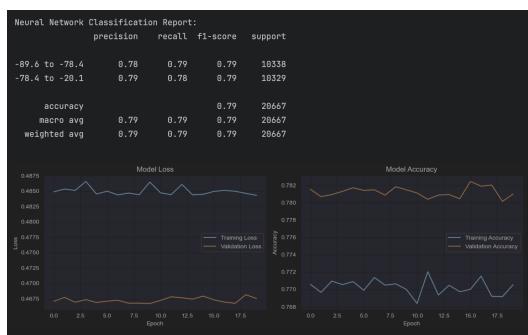
**Figure A25.** This figure shows the classification report for the Neural Network for 3 bins separated for Dayside Southern OCB magnetic latitude. Accuracy of 0.62, but precision and recall for -76.8 to -20.1 bin is 0.71 and 0.74.

South Hemisphere Results:				
	Model Accuracy	Macro Avg F1	Weighted Avg F1	
0 Logistic Regression	0.580733	0.570631	0.570574	
1 Random Forest	0.626700	0.622589	0.622533	
2 KNN	0.578749	0.576389	0.576332	
3 Decision Tree	0.609232	0.606344	0.606287	
4 Neural Network	0.624135	0.621575	0.621520	

**Figure A26.** Classifier Models for Dayside Southern hemisphere with 3 evenly cut bins for the magnetic latitudes. Random Forest and Neural Network have highest scores ( 0.63).

North Hemisphere Results:				
	Model Accuracy	Macro Avg F1	Weighted Avg F1	
0 Logistic Regression	0.755486	0.753961	0.753951	
1 Random Forest	0.789265	0.789265	0.789265	
2 KNN	0.763817	0.763745	0.763743	
3 Decision Tree	0.773332	0.773308	0.773307	
4 Neural Network	0.785350	0.785348	0.785349	

**Figure A27.** Classifier Models for Dayside Northern hemisphere with 2 evenly cut bins for the magnetic latitudes. Random Forest and Neural Network have highest scores ( 0.79).



**Figure A28.** This figure shows the classification report for the Neural Network for 2 bins separated for Dayside Southern OCB magnetic latitude. Accuracy of 0.79, but precision and recall for -76.8 to -20.1 bin is 0.79 and 0.79.

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.746746	0.745613	0.745621
1	Random Forest	0.783616	0.783605	0.783606
2	KNN	0.750665	0.750647	0.750648
3	Decision Tree	0.768713	0.768604	0.768606
4	Neural Network	0.785987	0.785984	0.785984

**Figure A29.** Classifier Models for Dayside Southern hemisphere with 2 evenly cut bins for the magnetic latitudes. Random Forest and Neural Network have highest scores ( 0.78).

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.519504	0.500913	0.501009
1	Random Forest	0.663484	0.659601	0.659674
2	KNN	0.610445	0.608477	0.608550
3	Decision Tree	0.636136	0.631288	0.631361
4	Neural Network	0.653967	0.651467	0.651535

**Figure A30.** Classifier Models for Nightside Southern hemisphere with 3 evenly cut bins for the magnetic latitudes. Random Forest and Network have highest scores ( 0.66).

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.543819	0.520154	0.520070
1	Random Forest	0.623397	0.619989	0.619932
2	KNN	0.567942	0.565907	0.565848
3	Decision Tree	0.604722	0.602030	0.601977
4	Neural Network	0.611874	0.607097	0.607038

**Figure A31.** Classifier Models for Nightside Northern hemisphere with 3 evenly cut bins for the magnetic latitudes. Random Forest and Network have highest scores ( 0.62).

South Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.303647	0.270210	0.270124
1	Random Forest	0.464127	0.457392	0.457391
2	KNN	0.409529	0.408654	0.408641
3	Decision Tree	0.434668	0.430041	0.430052
4	Neural Network	0.445874	0.436582	0.436608

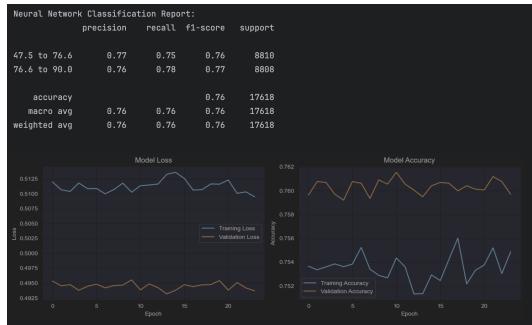
**Figure A32.** Classifier Models for Nightside Southern hemisphere with 5 evenly cut bins for the magnetic latitudes. Random Forest and Network have highest scores ( 0.46).

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.359235	0.323892	0.323897
1	Random Forest	0.440686	0.436689	0.436705
2	KNN	0.386707	0.385462	0.385470
3	Decision Tree	0.433307	0.430119	0.430142
4	Neural Network	0.434612	0.429136	0.429155

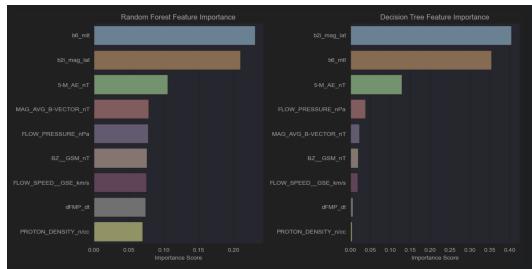
**Figure A33.** Classifier Models for Nightside Northern hemisphere with 5 evenly cut bins for the magnetic latitudes. Random Forest and Network have highest scores ( 0.44).

North Hemisphere Results:				
	Model	Accuracy	Macro Avg F1	Weighted Avg F1
0	Logistic Regression	0.710523	0.708444	0.708441
1	Random Forest	0.775003	0.774991	0.774991
2	KNN	0.734987	0.734941	0.734940
3	Decision Tree	0.760018	0.759923	0.759923
4	Neural Network	0.762345	0.762302	0.762302

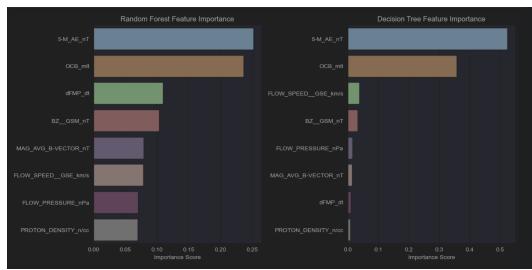
**Figure A34.** Classifier Models for Nightside Northern hemisphere with 2 evenly cut bins for the magnetic latitudes. Random Forest and Network have highest scores ( 0.78).



**Figure A35.** This figure shows the classification report for the Neural Network for 2 bins separated for Nightside Northern OCB magnetic latitude. Accuracy of 0.79, but precision and recall for 76.6 to 90.0 bin is 0.76 and 0.78.



**Figure A36.** Shows the Feature importance for separating the 2 bins in Nightside Northern OCB data, with OCB mlt gaining highest importance in Random Forest.



**Figure A37.** Shows the Feature importance for separating the 2 bins in Dayside Northern OCB data, with 5-M AE (nT) gaining highest importance in Random Forest and Decision Tree.