

# Using neural networks to study solar wind interactions in Earth's magnetic field.

Arnout Coenegrachts, Lukas Vinoelst  
Supervised by Sara Jamal

April 2021

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Magnetospheric Multiscale Spacecraft - MMS . . . . .	2
2.2	Exploited data in the study . . . . .	3
<b>3</b>	<b>Neural Networks</b>	<b>5</b>
3.1	General concept . . . . .	5
3.2	Multilayer Perceptron - MLP . . . . .	6
3.3	Fully Convolutional Network - FCN . . . . .	7
<b>4</b>	<b>Discussion and results</b>	<b>8</b>
4.1	Technical specifications . . . . .	8
4.2	MLP model . . . . .	8
4.3	FCN model . . . . .	11
<b>5</b>	<b>Conclusions</b>	<b>11</b>
5.1	Summary and interpretation of the results . . . . .	11
5.2	Possible future prospects . . . . .	12

## 1 Introduction

Nowadays, it is near impossible to avoid the widely-popular topics of Machine Learning (ML) and Deep Learning (DL) when surveying recent publications in the field of astrophysics. These concepts rely on the ability of systems to identify patterns in data and make decisions based on these detected characteristics. In this way the system is able to analyse the data in an automated way, barely requiring human input. A large variety of these learning techniques is being explored and refined to carry out analysis on astrophysical observations as well as simulations. Among these techniques, are supervised methods such as Neural Networks (NN) which employ tagged training data to achieve an analytical model, and unsupervised methods as clustering algorithms that extract information without any *a-priori* knowledge about the data (i.e., labels). An example of the application of clustering techniques to the detection of magnetic reconnection events in the Earth's magnetic field using clustering particle distribution data is discussed in Dupuis et al. (2020) [1]. Another example is the prediction of Solar Flare classes by applying various machine learning techniques to UV brightening and Magnetograms by Nishizuka et al. (2017) [2]. Using supervised Neural Networks, rather than unsupervised methods, Breuillard et al. (2020) [3] showed promising results for the classification of traversed regions of near Earth space from satellite data. We use the latter reference paper and supplementary material as a basis of the exploration done in this report. Similarly to that paper, classification of plasma regions employing convolutional networks on MMS energy distributions was investigated by Olshevsky et

al. (2019) [4], this application featured a 3D network rather than the more common 2D implementation. Yet another example of the use of AI in spaceweather is the CACTus program discussed during the lectures, which uses Hough transform techniques to detect the occurrence of Coronal Mass Ejections in LASCO images of the Sun. Reference is made to the work by Robbrecht et al. (2004) [5]. Another recent example of Neural Networks used in Astronomy research is the classification of galaxy morphology, described in Tadaki et al. (2020) [6]. Returning to the spaceweather domain, Camporeale et al. (2017) [7] proposed a Machine Learning classification algorithm for the origin of Solar Winds based on a technique called Gaussian Process. Convolutional Neural Networks find applications in countless other research fields as well. Examples include imaging and prediction applications in the medical world as presented by Milletari et al. (2016) [8], or crowd segmentation as presented by Kang et al. (2014) [9].

In Breuillard et al 2020 [3], the work focused on classification tasks, where the machine/deep learning techniques are employed to identify patterns and classify the data into specific regions categories. Following this reference work, classification tasks will be explored in this report. We will employ two distinct Neural Network architectures, a Multilayer Perceptron (MLP) (Gardner et al. (1998) [10]) and a Fully Convolutional Network (FCN) (Milletari et al. (2016) [8], Kang et al. (2014) [9], Long et al. (2014) [11]), to classify regions traversed by a renowned heliophysics NASA satellite mission known as Magnetospheric Multiscale Spacecraft (MMS) mission, and, from obtained results, analyse the misclassification and assess the performances of the NNs.

## 2 Data

### 2.1 Magnetospheric Multiscale Spacecraft - MMS

Interactions of Earth's magnetic field with Solar winds is a widely studied topic in the space weather community. Among the renowned legacy missions in this field is the Magnetospheric Multiscale Spacecraft (MMS) mission launched in March 2015, in which four probes were arranged in a tetrahedral formation along highly elliptical geocentric orbits, in order to spend significant time in specific regions of interest. The main goal of the two mission phases, dayside magnetopause regions and nightside magnetotail regions, is the collection of in-situ data and the discovery of interesting insights of magnetic reconnection processes, Burch et al. (2016)[12] In MMS, magnetic and electric field data as well as particle distribution data is collected at impressive cadences. The particle distribution data for instance is observed at a temporal resolution of 30ms for electrons and 150ms for ions, Pollock et al. (2016)[13]. These resolutions are necessary because the Electron Diffusion Regions (EDR) associated with magnetic reconnection events are both small and fast-moving and thus can easily be missed in the case of low measurement cadences.

A large amount of data (approximately 100 gigabits per day) is continuously being collected by the MMS mission [3]. However, transmission capabilities only allow for about 4 Gigabits of high resolution data to be sent to Earth per day, which present a very important problematic: "What data to transfer?". The method currently in use is the Scientist-In-The-Loop (SITL) method that consists in the survey data being reviewed by a human expert who decides whether (or not) to transfer the high resolution data to Earth's stations. With the ongoing development of Machine Learning methods and techniques, automated options become available to explore, such as Neural Networks or unsupervised algorithms. The idea of applying ML techniques to the data as a substitute for the SITL review, is among the possibilities that are unlocked by the ever-increasing ML research field. Using ML, a direct application would be to classify the distinct regions traversed by the spacecraft, as well as detecting EDR's and therefor magnetic reconnection events. Classified regions would consist of a interesting criteria to decide whether or not send the data from the satellite to Earth's stations. The use of ML would represent a vast improvement in efficiency given the sheer amount of data to be analysed.

In this report, we study the classification of MMS data using Neural Networks, following the work by Breuillard et al. (2020)[3]. Two architectures of neural networks are constructed and applied to MMS data with the aim to classify the traversed regions of Earth's magnetic field based on the observed field and particle data. The achieved classification, as well as misclassification, results provide insight to the NN performances and the characteristics of the physical processes occurring in these regions.

## 2.2 Exploited data in the study

Table 1: Features labels and corresponding descriptions.

Feature label	Feature description
$B_x$	$x$ -component of magnetic field
$B_y$	$y$ -component of magnetic field
$B_z$	$z$ -component of magnetic field
$B_{tot}$	Magnitude of magnetic field
$n$	Ion density
$V_x$	$x$ -component of ion bulk velocity
$V_y$	$y$ -component of ion bulk velocity
$V_z$	$z$ -component of ion bulk velocity
$V_{tot}$	Magnitude of bulk ion velocity
$T_{\parallel}$	Parallel ion temperature
$T_{\perp}$	Perpendicular ion temperature
$T_t$	Total ion temperature

Table 2: Region labels and full description.

Region label	Region name
SW	Solar wind
FS	Ion foreshock
BS	Bow shock
MSH	Magnetosheath
MP	Magnetopause
BL	Boundary layer
MSP	Magnetosphere
PS	Plasma sheet
PSBL	Plasma sheet boundary layer
LOBE	Lobe

The exploited MMS data in this study corresponds to the satellite observations obtained during the period of 2016-2019. The collected data is downloaded using the AIDAPy<sup>1</sup> open-source software and is resampled to the FPI instrument cadence of 4.5 s [3]. This total dataset consists of 1,331,133 time frames of 12 features as listed in Table 1. These features are given in the Geocentric Solar Ecliptic (GSE) coordinates of the MMS. The time frames have been labeled with regions in the Earth's magnetic field, where these labels are inferred from the SITL reports and supplemented for underpopulated regions. These supplemented labels are made based on physical properties observed from the features (e.g., magnetic field components, etc.). Training (56.25%), test (25%) and validation (18.75%) sets are randomly generated from the full dataset. Depending on the employed neural network architecture, the measured time frames are either considered as is, or are divided into blocks of 40 frames corresponding to intervals of 3 minutes. This 40-frame length in Breuillard et al. (2020)[3] was chosen because it minimizes the amount of padding the MMS timeframes need, while also preserving a sufficient number of datapoints in each block. Table 3 shows how many time series occurrences each region has. Following Breuillard et al. (2020) [3], the input data to the FCN model is standardized.

Table 3: Number of time series occurrences for each region.

Region	Time series 2016-2018	Time series 2019	Total time series
SW	1568	562	2130
FS	3209	1505	4714
BS	1646	682	2328
MSH	3144	1130	4274
MP	2943	821	3764
BL	3349	1195	4544
MSP	2714	1944	4658
PS	2451	1125	3576
PSBL	2216	487	2703
LOBE	964	504	1468
total	24204	9955	34159

Figures 1 show the time series data for the 2016-2018 and 2019 periods. This data converted into histograms shows the distribution of the different features in each of the plasma regions, cf. Figure 2. For example, it is conceivable that the Solar Wind has low  $B_i$  components, given that it refers to the furthest region from the

<sup>1</sup><https://gitlab.com/aidaspace/aidapy>

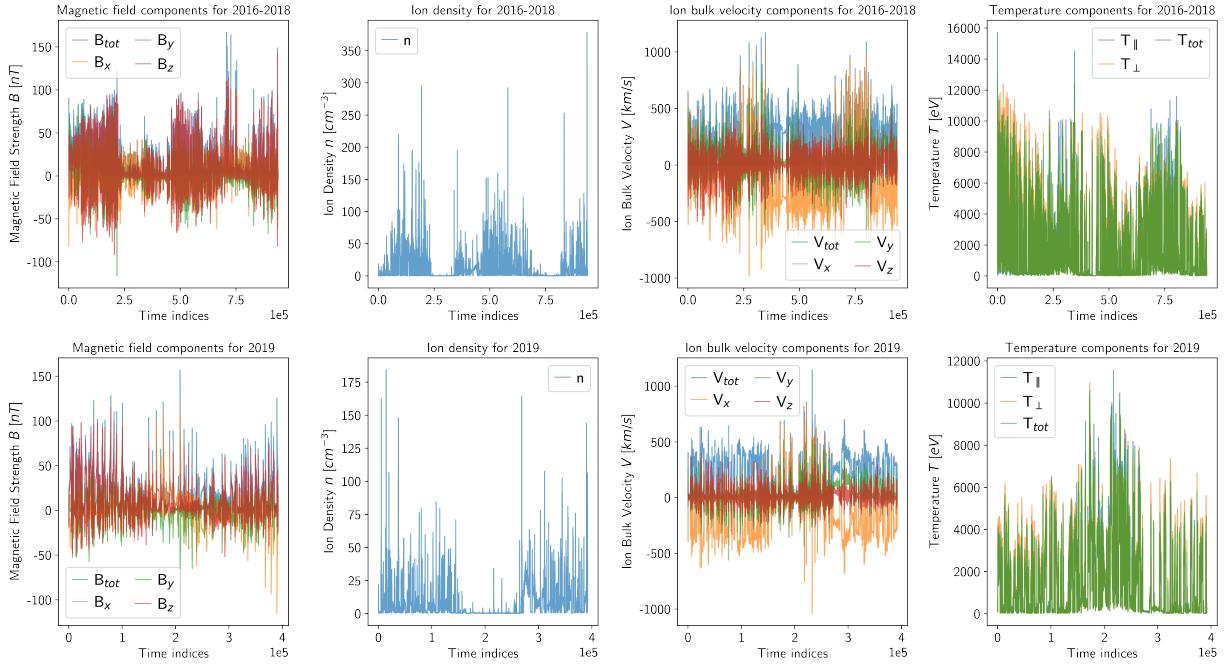


Figure 1: Time series data for 2016-2018 (top) and 2019 (bottom). Plotted data are, from left to right, the magnetic field components, the ion density, the bulk ion velocity components and the temperature components.

magnetic field of the earth and that it has a large negative  $V_x$  component, due to material coming from the Sun towards the Earth. There are some regions that are more interesting than others from a science point of view. For example, it is of great scientific interest to recognize patterns associated to the Magnetopause and the Plasma sheet, given that these two regions are the Near-Earth hosts where magnetic reconnection largely occurs. The ability to recognize interesting patterns and classify them correctly would allow to detect whether a spacecraft is traversing these regions or not. As stated, the interesting detected patterns could be linked to physical processes such as reconnection and might provide opportunities to automatize data processing pipelines and solve the transmission problem presented in earlier sections. Among interesting features of the Magnetopause region seen in the histograms in Figure 2, we can cite the multimodality (i.e., multiple peaks in the distribution/histograms) in the  $V_y$  component distribution, the left-skewed distribution of the  $V_x$  component, the distribution of  $V_{tot}$  and the magnetic field components in which  $B_z$  displays a right-skewed shifting towards a centered multimodal distribution. The magnetic field distributions and velocity distributions connected to the plasma sheet seem very similar to those of the lobe and plasma sheet boundary layer. The most notable differences lie with the  $B_x$  component which is more centralized around 0 for the plasma sheet as well as the density and temperature distributions, which displays different shapes than the plasma sheet boundary layer and lobe region altogether. Other features, such as multimodal distributions in the Magnetosheath, are visible as well suggesting the presence of different particle populations. The histograms of the data also show that the densities associated to Solar wind, the shock regions and the Magnetosheath are notably larger than the other regions, which is to be expected as these are made up of Solar Wind plasma, with densities ranging from 3 to up to  $10 \text{ ions/cm}^3$  depending on the speed/type of Solar Wind. Also notable is the higher temperature in the FS, BS, and MSH regions than the SW region, due to the plasma passing through the shock. When inspecting Fig. 1, periods of low density and magnetic field associated to the Magnetopause, Magnetosphere, Plasma sheet and Lobe regions (MP, BL, MSP, PS, PSBL, lobe) can be identified.

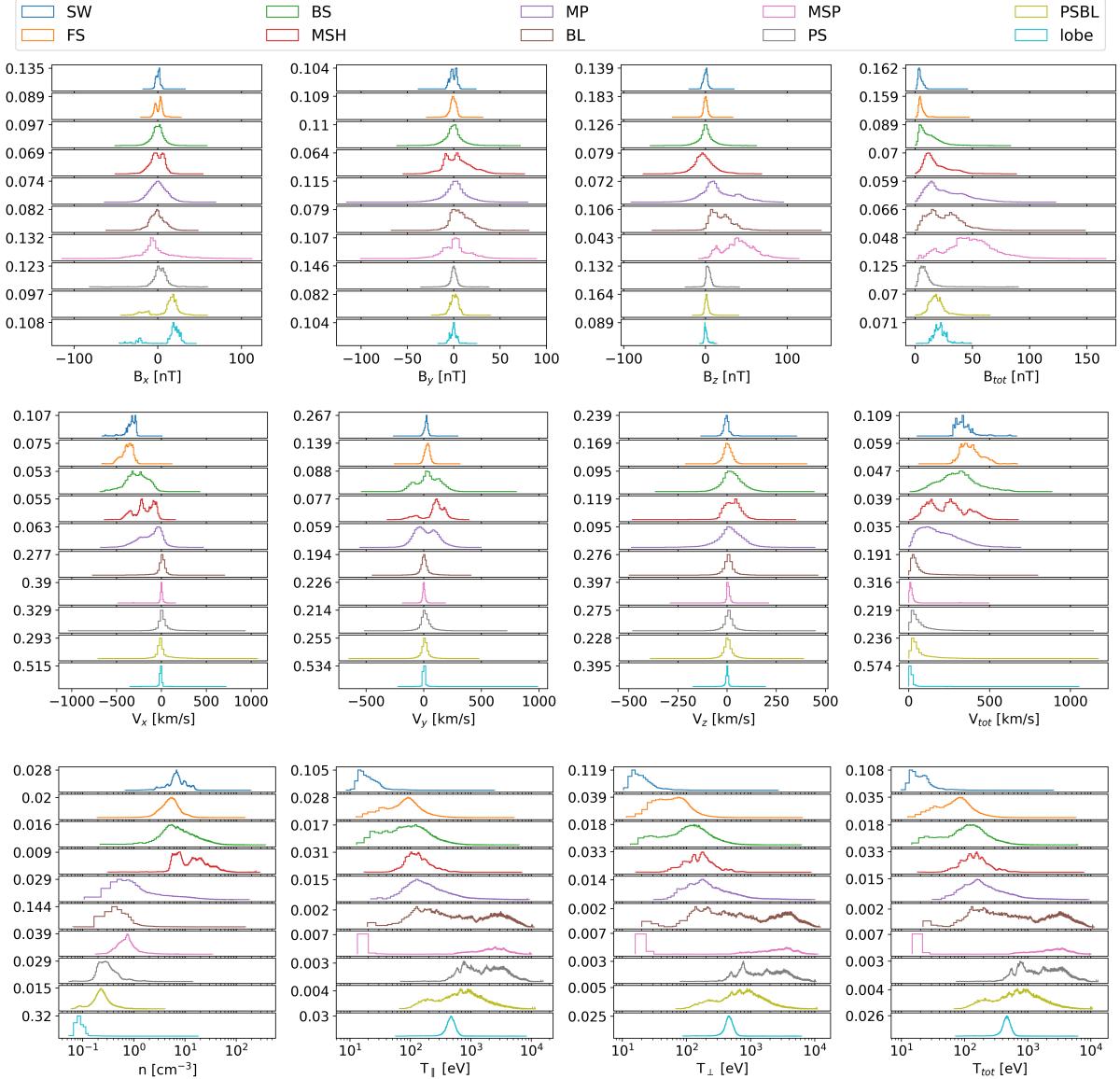


Figure 2: Normalised histograms showing the distribution of the total magnetic field  $B_{\text{tot}}$  and its components  $B_i$ , the total velocity  $V_{\text{tot}}$  and its components  $V_i$ , the particle density  $n$  and the total temperature  $T_{\text{tot}}$  and its parallel and perpendicular components  $T_{\parallel}$  and  $T_{\perp}$ , for each of the different plasma regions.

### 3 Neural Networks

#### 3.1 General concept

The NN original concept is based on the functioning of a real biological set of neurons in a human brain. The architecture of a Neural Network (NN) allows to process data through layers of connected nodes, referred to as "neurons", that works as a collective system and generate an output by applying an "activation function" to the input data. To add non-linearity to this model, the activation function can be non-linear. For a complete and in-depth treatment of Machine Learning concepts and algorithms, one should consult the book "Pattern Recognition and Machine Learning" by Bishop (2007) [14]. Another valuable source is the Deep Learning article by LeCun et al. (2015) [15] which gives detailed information on general supervised learning

concepts and focuses on convolutional and recurring Neural Networks. Despite the large variety of different kinds of neural networks, this report studies two architectures, following the reference work of Breuillard et al 2020 [3]: the "MLP" network and the "FCN" model. In the next subsections, we further develop the specificity of these two architectures.

### 3.2 Multilayer Perceptron - MLP

Perhaps the most straightforward example of a classification neural network is the Multilayer Perceptron (MLP). This network consists of the input and output layers, with hidden layers called "perceptrons" in the hidden layers. The perceptron layers consist of nodes applying an activation function to the inputs, commonly the "rectified linear unit" (relu) function that sets negative inputs to zero and only retains positive input values.

The specific configuration of the MLP model used in this report is illustrated in Fig. 3. The first layer of the network refers to the input layer, which processes the data, i.e. a vector of 12 components representing the data features listed in Table 1. Following the input layer, the data is processed through a succession of dense layers consisting of 128 nodes and followed each by: a batch normalization layer, a dropout layer (which sets to 0 a fraction of the data, 1% in this case), and an activation layer of the relu activation function. The final layer of the MLP architecture of this study corresponds to a dense layer to compute the output labels in the form of a categorical vector of 10 components representing the scores associated to the 10 different regions listed in Table 2. The last output layer is activated by a "softmax" activation function. The batch normalization standardizes the data before feeding it to the next layer, which enhances computation time in the sense that fewer epochs will be needed to achieve a certain precision.

The activation function serves the purpose of providing a non-linearity to the nodes. This non-linearity is a key concept to reproduce the behaviours of neural networks. The dropout layer even provides an additional non-linearity.

The MLP model is trained for 100 epochs, with batches of 64, employing an early termination method based on the evolution of the validation loss over time. Via the monitoring of the validation loss, the training is terminated when a stagnation is detected in order to avoid overfitting. The "Adam" optimizer is exploited with a learning rate of  $1 \cdot 10^{-4}$  and the loss function is chosen as the categorical cross-entropy. Training was terminated at epoch 55 by the early stopping. The evolution of the loss and the accuracy for the training and validation sets are given in Fig. 4.

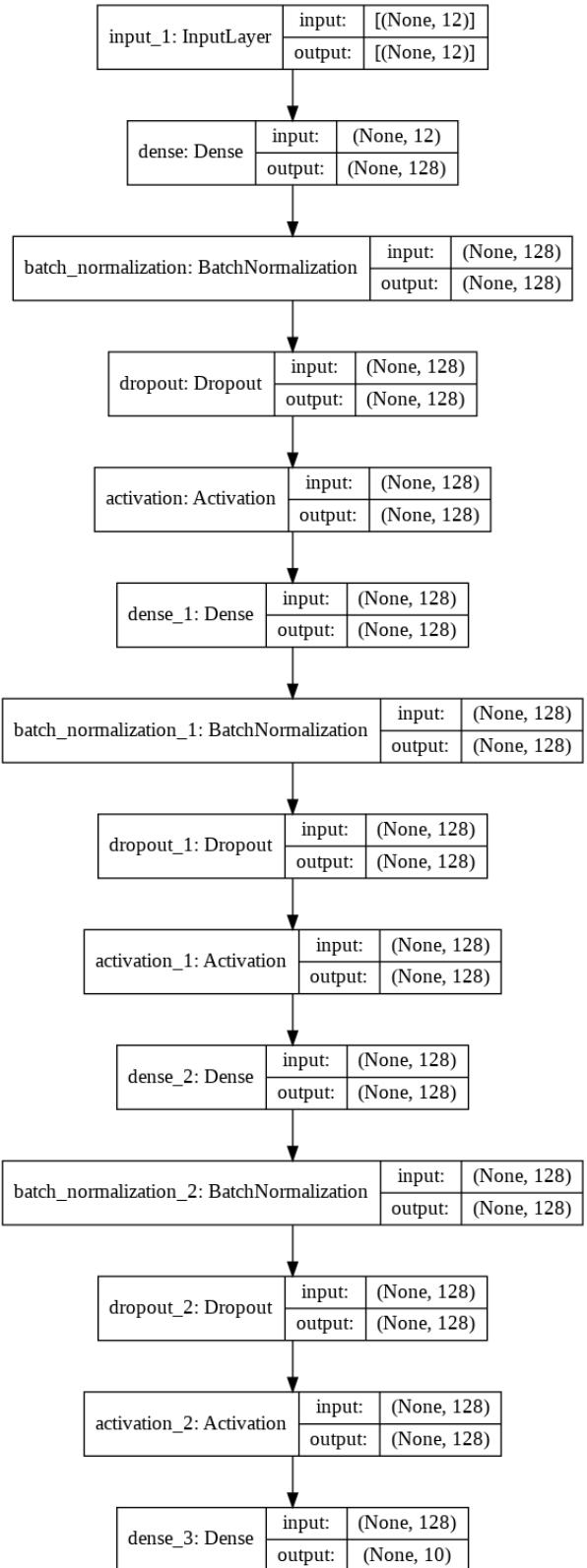


Figure 3: MLP model layer architecture.

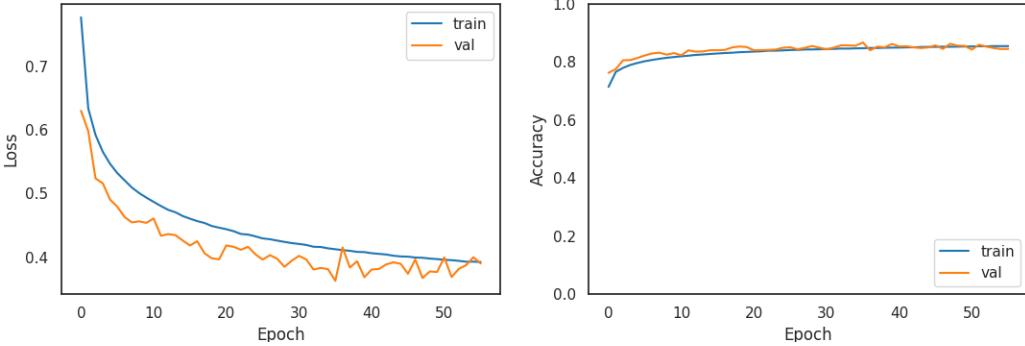


Figure 4: Training (blue) and validation (orange) losses (left) and accuracies (right) during the training phase of the MLP model.

### 3.3 Fully Convolutional Network - FCN

A drawback of using the MLP model is that temporal information (dates) of the data are disregarded, given that the features are considered as static timeframes, disconnected from each other.

An improvement would be to use a Fully Convolutional Network (FCN), which considers timeseries in blocks of 40 timeframes and convolves them with specific filters to detect features.

An overview of the used FCN model configuration is given in Fig. 5.

The FCN model consists of 11 layers total, the first of which is an input layer. This layer accepts three-dimensional input data with shape  $(n, 40, 12)$ , where  $n$  refers to the number of observations that are dispatched into 40-length for each of the 12 features, as listed in Table 1. After this layer, there follows a sequence of three time convolution blocks. These time convolution blocks consist each of a convolution layer Conv1D, batch normalisation layer and an activation layer. Each convolution layer produces a convolution between the input and an amount of filters to get temporal features from the input.

The first convolution layer has 256 filters and a kernel size of 8, the second convolution layer has 512 filters and a kernel size of 5 and the third convolution layer has 256 filters and a kernel size of 3.

As previously stated, the batch normalization layer normalizes the output from the preceding convolution layer, after which the activation layer applies the relu activation function to the output from the batch normalization layer. The output of the third convolution block will pass through a global average pooling layer and a dense layer to finally give the results of the classification.

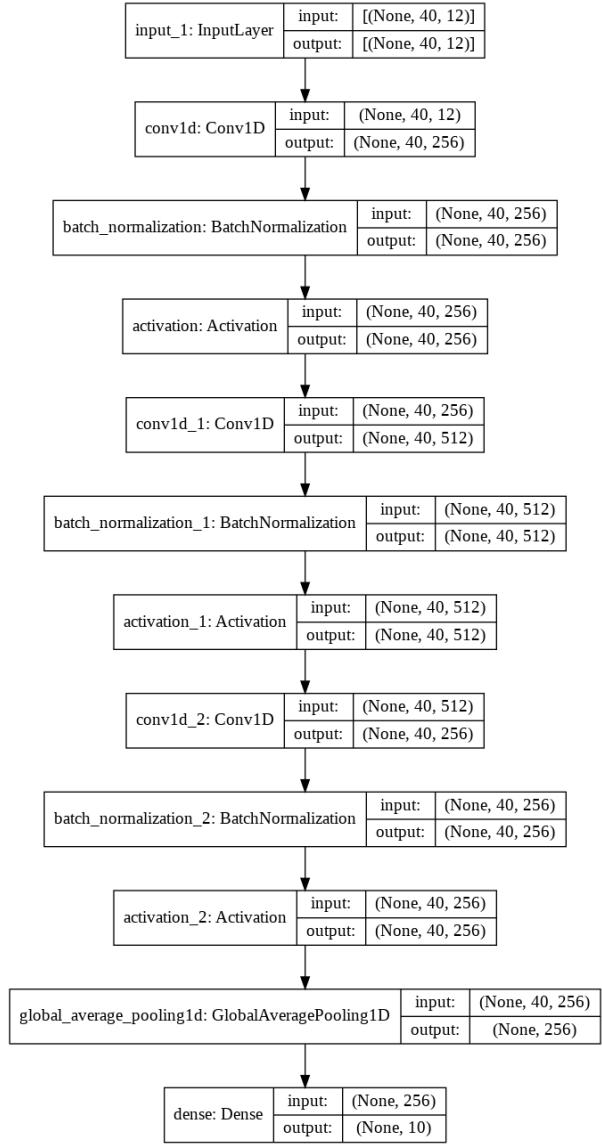


Figure 5: FCN model layer architecture.

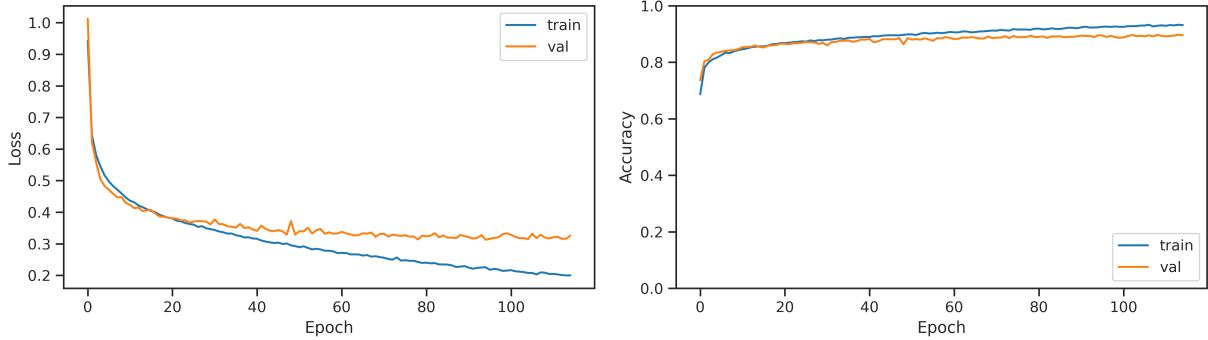


Figure 6: Training (blue) and validation (orange) losses (left) and accuracies (right) during the training phase of the FCN model.

Following The work by Breuillard et al 2020 [3], the FCN model was trained in batches of size 92, employing the same early termination method used for the MLP method. The training was terminated at epoch 115 by early stopping. This model also used the "Adam" optimizer with a learning rate of  $2 \cdot 10^{-5}$  and the categorical cross-entropy as the loss function. Figure 6 shows the loss and accuracy of the classification of the test data.

## 4 Discussion and results

### 4.1 Technical specifications

At first, the models were trained on the local machines of the project authors (LV and AC). However, the training time on local machines for the FCN training was underperforming. We chose to execute the full project (MLP and FCN) using **Google Colaboratory**, where GPU can be used rather than just CPU. This change drastically improved the training time for the FCN model. The early stopping termination for the MLP model on **Google Colaboratory** was epoch\_55 instead of epoch\_85 locally, however we noticed a small increase of the runtime per epoch on **Google Colaboratory** in comparison with the local machine Intel® Core™ i5-1035G1 CPU @ 1.00GHz x 8 processor). We thoroughly investigated the origin of this speed decrease and found that it related mainly to **Python** packages versions as well as the nature of GPU limits on **Google Colaboratory**. We set a coding environment locally similar to **Google Colaboratory** and attained the exact runtime with a different early stopping. Our investigation concluded that performances on **Google Colaboratory** are satisfactory and that **Python** packages requires to set random seed in order to preserve reproducibility of our study results.

Another important argument to move the process over to **Google Colaboratory** was the interest of having a shared environment to compare the results between the FCN and the MLP. The training time for the FCN model was 37 seconds for the first epoch, and 3 seconds per epoch afterwards, for a total training time of 379 seconds (6 minutes 19 seconds). The training time for the MLP model was 45 seconds per epoch on average for a total of 55 epochs, resulting in a total training time of 41 minutes 15 seconds.

For reproducibility and comparability of the results, the random seeds of the **NumPy**-, **keras**- and **Tensorflow**-packages were fixed to 42.

### 4.2 MLP model

On the test set, the MLP model achieves overall satisfactory performances in classifying of the data as illustrated by the confusion matrices in Fig. 7a. However, some regions stand out as not being well classified by the model, such as the Solar Wind labels with only 38% of true positives and the Bow Shock with a 58% correct classification rate. Another region with a relatively low true predictions (77%) is the Magnetopause.

The Solar Wind & Bow Shock misclassifications may be related to the fact that these very dynamical regions are affected from the misclassifications obtained on the adjacent regions such as the Ion foreshock and the Magnetosheath. Hence, very similar physical processes and/or particle populations might occur in these regions, which inevitably leads to similar patterns that are mistakenly identified by the NN model. The Magnetopause corresponds to the thin boundary layer between the Magnetosphere and the Magnetosheath. Thus, it can very likely features particle populations associated with both of these neighbouring regions. Indeed, Fig. 7a shows that the Magnetopause leans towards the associated labels of the Bow Shock (BS), the Magnetosphere (MSP), the Boundary Layer (BL) and the Magnetosheath (MSH).

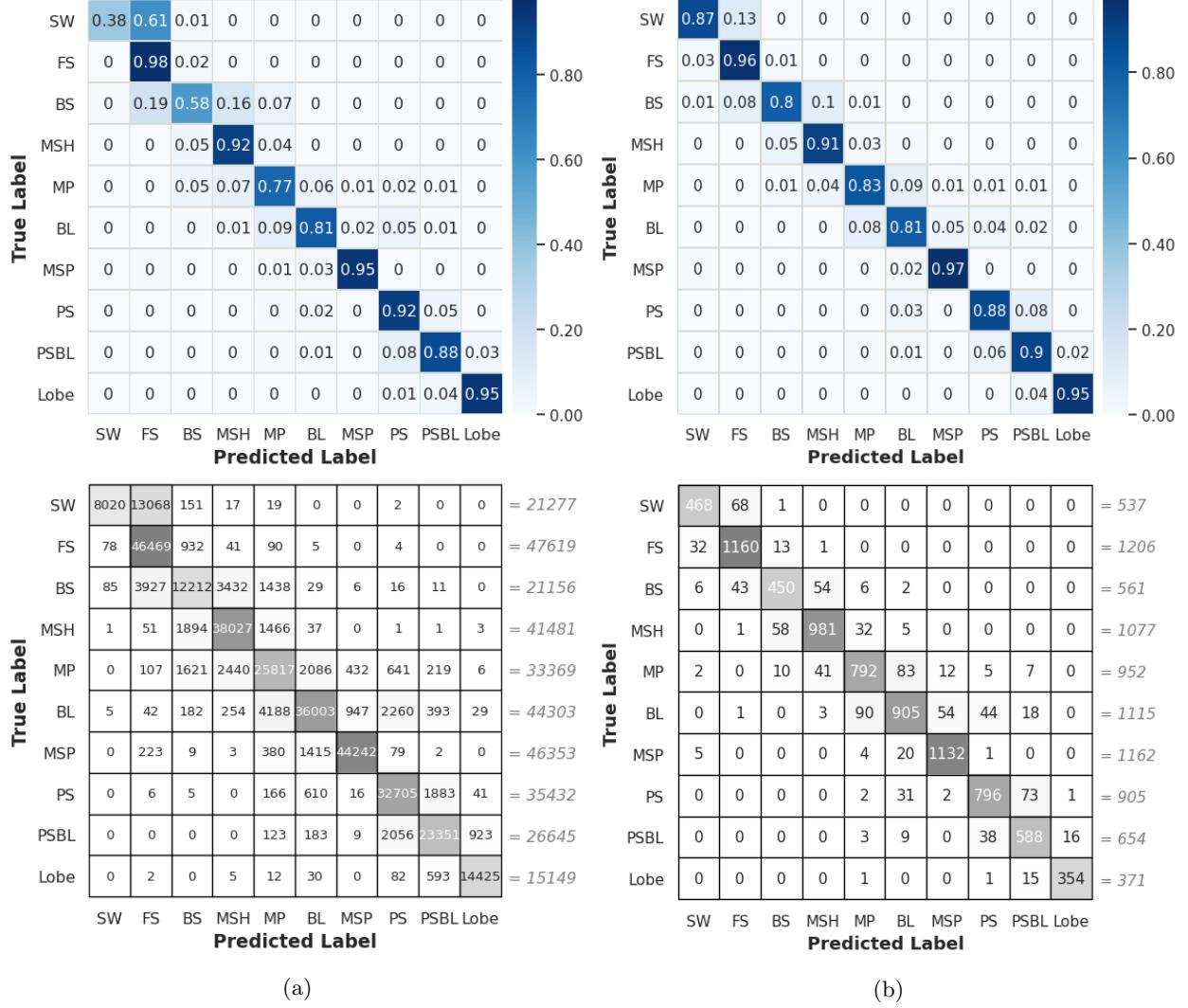


Figure 7: Confusion matrices for the MLP model (a) and the FCN model (b).

To evaluate how well the model performs, the classification metrics, from the `sklearn` Python library, is used. We report our results in Table 4 for the MLP model. This report gives the precision, recall, f1-score and the accuracy of the predictions. For reference, the classification metrics are defined as follows (also used in Breuillard et al. [3]):

- The precision is the proportion of correct detections.

$$precision = \frac{TP}{TP + FP} \quad (1)$$

Here,  $TP$  is the amount of True Positives, while  $FP$  is the amount of False Positives.

- The recall is the percentage of detected positives, also called the True Positive Rate (TPR).

$$recall = \frac{TP}{TP + FN} \quad (2)$$

Here,  $FN$  is the amount of False Negatives.

- The F1-score is defined to be the harmonic mean of the precision and the recall. Typically, a high precision means a lower recall and vice versa, so the F1-score show the trade-off between these quantities.

$$F1\text{-score} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (3)$$

- Finally, the accuracy is given by

$$accuracy = \frac{TP}{N_p} \quad (4)$$

Here,  $N_p$  is the total amount of predictions.

The report also features the (macro) average and weighted averages of each of these quantities across all regions/labels. Table 4 confirms that the classification of the Solar Wind (SW) and Bow Shock (BS) attains very poor recall and f1-scores indicating a low detection rate. This is also true for the BS region. Notable is that for the Magnetopause (MP) and Plasma Sheet Boundary Layer (PSBL) regions, precision, recall and f1-score attain similar numbers. Overall, the model attains metrics around 85% From our results, the misclassification of small and interacting regions requires further improvement.

An additional statistical tool is the Receiver Operating Characteristic (ROC) curve, that represents the True Positive Rate  $TPR$  (recall) against the False Positive Rate  $FPR$ , for different threshold values. The  $FPR$  is defined as follows.

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

with  $TN$  the amount of True Negatives.

The threshold value refers to the criterion that determines whether an element is classified as a "negative" or as a "positive" label. At a threshold of zero, all elements are classified as "positives" ( $TPR = FPR = 1$ ), while at a threshold of one, they are all classified as "negatives" ( $TPR = FPR = 0$ ). A perfect classifier would be located at the top left corner, corresponding to  $TPR = 1$  and  $FPR = 0$ . By definition, the closer the ROC curve of a certain class comes to this point, the better the model is at classifying that particular class. The area underneath a ROC curve, called the AUC (Area Under Curve), gives the probability that the model predicts for a random positive instance a higher score than for a random negative instance. Generally speaking, a higher AUC-value (closer to one) means that the model is highly performant [16].

The ROC curves for the MLP model are reported in Fig. 8a. The Bow Showk (BS) and Magnetopause (MP) ROC curves stand out as the underperforming curves with respective AUC values of 0.97 and 0.98. However, from the ROC curves is not highly visible that the Solar Wind (SW) classification is underperforming. Therefore, an analysis should make use of the different statistical tools to assess the overall performances of the NN.

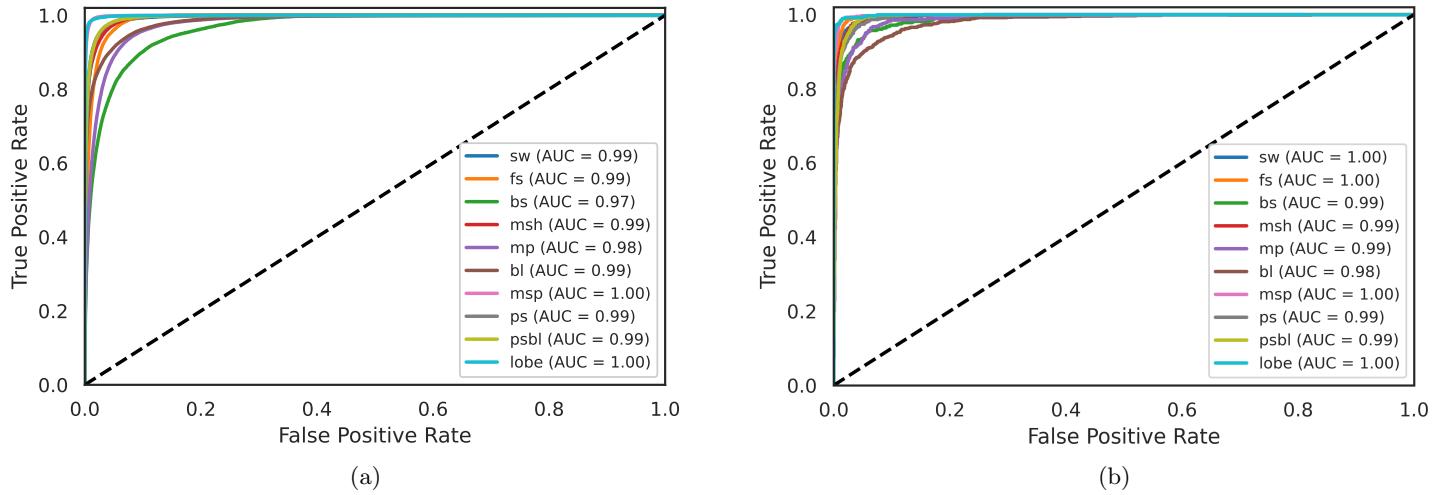


Figure 8: ROC curves for the MLP model (a) and FCN model (b). Plotted is the True Positive Rate versus the False Positive Rate. Lower curves experience more False Positives before reaching the True Positive Rate of 1. The AUC in the legend refers to the area under the curve.

### 4.3 FCN model

The confusion matrices given in Figure 7b show the classification of test data of the FCN model. The classifications is pretty good for all regions, with the worst one being the Bow Shock with only 80% correctly classified. Apart from few outliers, the vast majority of misclassifications occur in regions that lie close to each other. For example, the Magnetosheet (MSH) lies between the Bow Shock (BS) and the Magnetopause (MP), so some of their plasma properties are likely similar, which would make it harder for the model to distinguish between these regions.

Table 5 gives the classification report from `sklearn` for the FCN model. For most of the regions, the values of the quantities of our version of the model differ from the paper. The NN predicts labels for Solar Wind (SW), Fore Schock (FS), Magnetopause (MP), Boundary Layer (BL) and Plasma Sheet Boundary Layer (PSBL) regions with higher precisions and lower recall values using our compiled FCN model in comparison with the reference paper’s model [3], while the NN attains for the Magnetosphere (MSP) and Plasma Sheet (PS) a lower precision and higher recall in contrast. For the Bow Shock (BS), the Magnetosheet (MSH) and the Lobe (LOBE) regions, our FCN model attains simalr performances as Breuillard et al 2020. The accuracy and averages retain similar values in all cases (0.89).

Figure 8b shows the ROC curves for the FCN model and the AUC values for each of the regions. The displayed curves are nearly identical to the ones from the reference paper of Breuillard at al 2020, where the classification performances for the Boundary Layer (BL) are associated to the lowest AUC.

## 5 Conclusions

### 5.1 Summary and interpretation of the results

In this study, a Multi-Layer Perceptron (MLP) and a Fully-Connected Network (FCN) are constructed and trained to classify MMS mission data into plasma regions of Near-Earth space. Our study is based on the published work by Breuillard et al 2020 [3]. The two networks are trained and tested using labeled data from the Magnetospheric Multiscale Spacecraft (MMS) mission and achieved satisfactory results for the classification task. The FCN model was found to perform better overall but slightly less accurate when dealing with the thinnest boundary layers. When comparing the MLP and FCN model performances, it is very clear that the MLP underperforms when classifying the different regions, which implies that the

Table 4: The classification report for the MLP model

	precision	recall	f1-score
SW	0.98	0.38	0.54
FS	0.73	0.98	0.83
BS	0.72	0.58	0.64
MSH	0.86	0.92	0.89
MP	0.77	0.77	0.77
BL	0.89	0.81	0.85
MSP	0.97	0.95	0.96
PS	0.86	0.92	0.89
PSBL	0.88	0.88	0.88
LOBE	0.94	0.95	0.94
accuracy		0.85	
macro avg	0.86	0.81	0.82
weighted avg	0.86	0.85	0.84

Table 5: The classification report for the FCN model

	precision	recall	f1-score
SW	0.91	0.87	0.89
FS	0.91	0.96	0.94
BS	0.85	0.80	0.82
MSH	0.91	0.91	0.91
MP	0.85	0.83	0.84
BL	0.86	0.81	0.83
MSP	0.94	0.97	0.96
PS	0.90	0.88	0.89
PSBL	0.84	0.90	0.87
LOBE	0.95	0.95	0.95
accuracy			0.89
macro avg	0.89	0.89	0.89
weighted avg	0.89	0.89	0.89

convolution method is more effective to detect relevant temporal patterns in the physical timeseries and distinguish between the regions based on these patterns. The classification of regions improved vastly in the FCN model in comparison with the MLP architecture choice; the 38% classification rate was enhanced to 82% for the Solar Wind (SW) region and that of the Bow Shock (BS) region increased from 58% to 80%. The f1-scores also improved for most regions, despite 2 exceptions for the Boundary Layer (BL) and the Plasma Sheet Boundary Layer (PSBL) regions, suggesting that the MLP might be slightly better at resolving these thin boundary layers. A possible cause might be that the interesting patterns seen in these regions are resolved with a resolution that can be diluted by the convolutions performed in the FCN. The final FCN model achieved the best classification with a precision, recall and f1-score of 89%. The final MLP model achieved a classification with a precision of 86%, a recall of 85% and an f1-score of 84%. These ML techniques show promising results and may provide viable solutions to the problem of data selection in the context of spacecraft transmission limitations.

## 5.2 Possible future prospects

Several neural network architectures remain to be tested. Different configurations of similar models as tested in this report could lead to different classification performances. The availability of the large satellite datasets undeniably provides many opportunities to further investigate the use of ML and Deep Learning in this line of research, making the future of Spaceweather in this context look very promising.

## References

- [1] R. Dupuis, M. V. Goldman, D. L. Newman, J. Amaya, and G. Lapenta, “Characterizing magnetic reconnection regions using gaussian mixture models on particle velocity distributions,” *The Astrophysical Journal*, vol. 889, p. 22, Jan 2020.
- [2] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, S. Watari, and M. Ishii, “Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms,” *The Astrophysical Journal*, vol. 835, p. 156, Jan 2017.
- [3] H. Breuillard, R. Dupuis, A. Retino, O. Le Contel, J. Amaya, and G. Lapenta, “Automatic classification of plasma regions in near-earth space with supervised machine learning: Application to magnetospheric multi scale 2016–2019 observations,” *Frontiers in Astronomy and Space Sciences*, vol. 7, p. 55, 2020.
- [4] V. Olshevsky, Y. V. Khotyaintsev, A. Divin, G. L. Delzanno, S. Anderzen, P. Herman, S. W. D. Chien, L. Avanov, and S. Markidis, “Automated classification of plasma regions using 3d particle energy distribution,” 2019.

- [5] E. Robbrecht and D. Berghmans, “Automated recognition of coronal mass ejections (CMEs) in near-real-time data,” *Astronomy and Astrophysics*, vol. 425, pp. 1097–1106, Oct. 2004.
- [6] K.-i. Tadaki, M. Iye, H. Fukumoto, M. Hayashi, C. E. Rusu, R. Shimakawa, and T. Tosaki, “Spin parity of spiral galaxies II: a catalogue of 80k spiral galaxies using big data from the Subaru Hyper Suprime-Cam survey and deep learning,” *Monthly Notices of the Royal Astronomical Society*, vol. 496, pp. 4276–4286, 07 2020.
- [7] E. Camporeale, A. Carè, and J. E. Borovsky, “Classification of solar wind with machine learning,” *Journal of Geophysical Research: Space Physics*, vol. 122, no. 11, pp. 10,910–10,920, 2017.
- [8] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” 2016.
- [9] K. Kang and X. Wang, “Fully convolutional neural networks for crowd segmentation,” 2014.
- [10] M. Gardner and S. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” *Atmospheric Environment*, vol. 32, no. 14, pp. 2627–2636, 1998.
- [11] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.
- [12] J. Burch, T. Moore, R. Torbert, and et al., “Magnetospheric multiscale overview and science objectives.,” *Space Sci Rev*, vol. 199, pp. 5–21, 2016.
- [13] C. Pollock, T. Moore, A. Jacques, and et al., “Fast plasma investigation for magnetospheric multiscale.,” *Space Sci Rev*, vol. 199, pp. 331–406, 2016.
- [14] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 10 2007.
- [15] Y. LeCun, Y. Bengio, and H. Geoffrey, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 05 2015.
- [16] T. Fawcett, “An introduction to roc analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. ROC Analysis in Pattern Recognition.