

Using Self-Supervised Machine Learning to Identify Pulsed Ionospheric Flows in SuperDARN Observations

Aman Bhalla

University of Southampton, University Rd, Southampton SO17 1BJ

26 April 2025

ABSTRACT

The classification of ionospheric flows is critical for understanding large scale magnetospheric convection, a process driven by magnetic reconnection and flux transfer events. This study applies a self-supervised SimCLR contrastive learning framework to automate the identification of Pulsed Ionospheric Flows (PIFs) from SuperDARN backscattered power data. Using observations from the Hankasalmi radar, models trained with and without multi-head attention mechanisms achieve high top-1 accuracies (98-99%), though this reflects some model overfitting. DBSCAN clustering of learned representations reveal a moderately cohesive group (silhouette score ≈ 0.53), containing $\approx 30\%$ PIFs, though groundscatter remains prominent in the group. Future work should focus on diversifying augmentations and implementing multi-modal features to further improve model generalisation to PIFs.

1 INTRODUCTION

Understanding ionospheric dynamics is critical to comprehending larger scale magnetospheric convection, a key process responsible for phenomena such as auroral events and the formation of radiation belts (Axford 1969). The primary driver of magnetospheric convection is the solar wind, which interacts with the magnetosphere in a driving process known as magnetic reconnection (Dai et al. 2024; Baumjohann and Treumann 2012). Signatures of reconnection have been observed in-situ by spacecraft. Observations from the ISEE-1 and -2 spacecraft revealed a time dependent reconnection process, occurring in a highly variable (bursty) manner (Russell and Elphic 1978). Understanding the variability in these reconnection bursts can help to resolve how reconnection modulates magnetospheric convection.

In the ionosphere, the signatures of reconnection bursts result in Poleward Moving Auroral Forms (PMAFs) which are known as Pulsed Ionospheric Flows (PIFs) in ground-based radar data (Vorobjev et al. 1975; Fear et al. 2017). Subsequently, the identification of PIFs can prove significant in resolving the dynamics of bursty reconnection (Sandholt et al. 1986; Lockwood et al. 1989). Arrays of high frequency radars such as The Super Dual Auroral Radar Network (SuperDARN) have been used to detect PIFs in the ionosphere (Chisham et al. 2007). Studies have been conducted to analyse and classify observed PIFs from SuperDARN data to help answer questions about the dynamics of magnetospheric convection (Provan and Yeoman 1999; McWilliams et al. 2000; Fear et al. 2017). In these studies, PIFs have been manually identified in time-series data.

However, manual classification of PIFs is time-consuming and does not provide the volume of classifications needed to answer high level questions regarding magnetospheric dynamics. As manual identification proves challenging, automated approaches can be adopted. Machine Learning (ML) techniques have shown significant advancements in the classification of time series data (Fawaz et al. 2019).

ML techniques have been successfully applied to SuperDARN datasets. For example, Liu et al. (2019) used Artificial Neural Networks (ANNs) and regression models to estimate the cross polar cap potential using time series data from SuperDARN. Kunduri et al. (2022) used a neural network to successfully

distinguish between SuperDARN backscatter modes. These studies typically rely on supervised methods to analyse SuperDARN data. However, there are a limited amount of labelled PIFs in SuperDARN data, as such supervised approaches may not be suitable. Unsupervised methods have been used on SuperDARN data. For example, Riberio et al. (2011) used depth-first search clustering to identify ionospheric backscatter, however this method used boxcar filtering, which reduced spatial and temporal resolution of ionospheric features. Robb et al. (2018) used unsupervised clustering methods which maintained high data resolution, however they presented limitations in speed and separation of clusters.

An alternative to traditional unsupervised learning is self-supervised learning (SSL), a method which does not require labelled data as it learns pseudo-labels from the data. This project aims to use a self-supervised learning framework known as SimCLR (A Simple Framework for Contrastive Learning of Visual Representations). SimCLR is a contrastive learning framework first introduced by Chen et al. (2020). SimCLR computes similarity scores between data points which act as pseudo-labels for classification. It has shown success in auroral image classification (Johnson et al. 2021) and consistently boosts performance for time-series classification (Zhang et al. 2022). By applying SimCLR to SuperDARN observations this study aims to develop an alternative to manual classification, to increase the volume of identified PIFs and develop an accurate and faster procedure to identify PIFs. This study will evaluate the effectiveness of SimCLR using downstream DBSCAN clustering and understand the effects of using attention mechanisms on classification performance.

2 BACKGROUND

This section outlines the theoretical background behind the magnetospheric processes that result in PIFs, followed by an overview of how the SuperDARN radar operates and is used to observe PIFs. Finally, the section concludes with a description of the SimCLR framework and ML principles.

2.1 MAGNETIC CONVECTION AND IONOSPHERIC SIGNATURES

2.1.1 Magnetic Reconnection

Magnetospheric convection ultimately begins with the solar wind, an expanding plasma which streams radially out of the Sun (Baumjohann and Treumann 2012). As the solar wind is highly conductive, it drags out the solar magnetic field in its direction of travel, becoming "frozen in" whilst forming the Interplanetary Magnetic Field (IMF). Figure 1 outlines the topography of these dynamics.

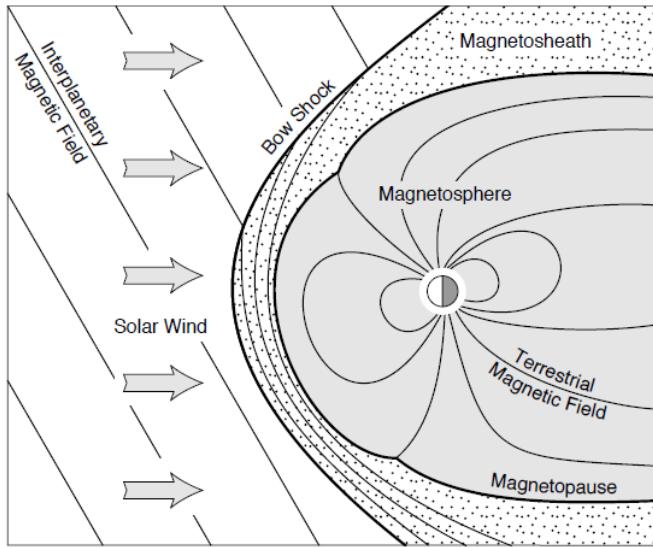


Figure 1: Illustration of the interaction between the IMF dragged out by the solar wind and its deflection around the Earth's magnetosphere. Ultimately, the solar wind stretches Earth's magnetosphere, forming a magnetotail at Earth's nightside (Baumjohann and Treumann 2012)

As the solar wind plasma encounters the terrestrial magnetic field it deflects around it (Russell et al. 2016). However, under certain conditions, solar wind plasma can enter the Earth's magnetosphere, the primary process responsible for this is magnetic reconnection (Dai et al. 2024).

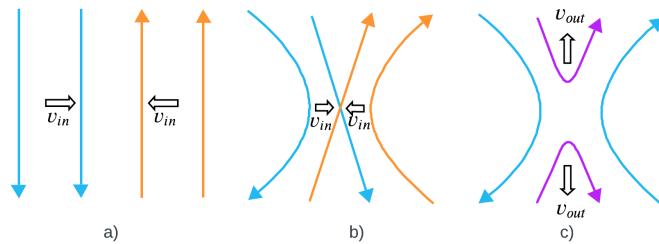


Figure 2: The evolution of the topology of antiparallel magnetic field lines during a magnetic reconnection event. a) the encounter of antiparallel field lines, b) the formation of an X-line and neutral point, c) reconfigured field line topology resulting in reconnection jets with mixed plasma populations (purple).

Reconnection occurs when an interplanetary field line oriented antiparallel to Earth's magnetic field encounters a field line at the dayside magnetopause. This topology is illustrated in part a) of Figure 2 by the red and blue antiparallel field lines of the IMF and magnetopause. The field lines are initially frozen into the plasma. As the field lines move towards each

other, the magnetic field vanishes at a local region known as a neutral point. At the neutral point, the plasma and magnetic field are no longer frozen in. This forms a topology known as an X-line illustrated in part b) of Figure 2. The antiparallel field lines from the IMF and magnetopause are then halved and merged at the neutral point, mixing both plasma populations (Baumjohann and Treumann 2012; Fear et al. 2017). This forms a new topology illustrated in part c) of Figure 2. As magnetic tension is released, plasma is ejected away from the neutral point, forming reconnection jets, transporting solar plasma into the magnetosphere (Baumjohann and Treumann 2012; Russell et al. 2016).

2.1.2 Flux Transfer Events

The signatures of reconnection events were first described by Russell and Elphic (1978) after in-situ observations from ISEE-1 and 2. When the spacecraft crossed the magnetosheath illustrated in Figure 1, a bipolar variation in the magnetic field component normal to the magnetopause was observed. This variation was spatially local and transient and became known as a Flux Transfer Event (FTE) and is the magnetospheric signature of reconnection bursts (Rijnbeek et al. 1982).

2.1.3 The Dungey Cycle and Poleward Moving Auroral Forms (PMAFs)

FTEs drive a convection process in the magnetosphere known as the Dungey cycle, first described by Dungey (1961), and visualised in Figure 3.

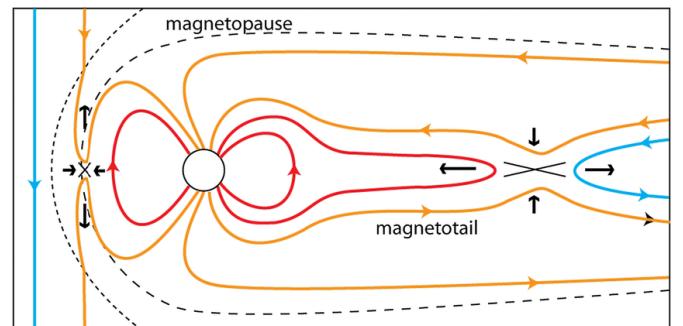


Figure 3: A southward IMF field line (blue) encounters magnetosphere field lines (orange) resulting in magnetospheric convection, with dayside and nightside reconnection indicated by X-lines (Eastwood et al. 2017).

The cycle occurs when a southward directed field line from the IMF encounters the dayside magnetopause and initiates a reconnection event, illustrated by the X line in Figure 3. As the field lines reconnect, they split open and the solar wind transports these field lines downtail (orange field lines in Figure 3) (Baumjohann and Treumann 2012). As the open halves of the field lines move down towards the magnetotail, they eventually meet at the nightside of the magnetosphere and reconnect again (Akasofu 1964).

The magnetospheric field lines have footprints on the Earth, as they move down tail, the motion of field lines drive signatures known as poleward moving auroral forms (PMAFs), first described by Vorobjev et al. (1975).

2.2 IONOSPHERIC OBSERVATIONS WITH SUPERDARN

PMAFs can be observed in the ionosphere using High Frequency (HF) coherent scatter radars such as the The Super Dual Auroral Radar Network ([SuperDARN](#)) ([Greenwald 1995](#)).

2.2.1 SuperDARN Radar Operation

SuperDARN is an array of ground based HF radars with a field of view covering extensive regions of the Northern and Southern hemisphere. Each radar consists of 16 antennas arranged in an array, transmitting radio waves into the ionosphere. The 16 different antennas (look directions) allow for the formation of a global view of plasma convection in the ionosphere, an example of this is displayed in Figure 4 ([Chisham et al. 2007](#)).

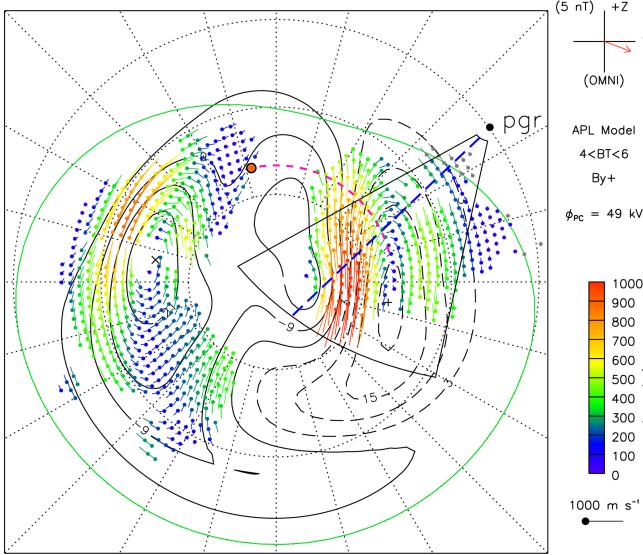


Figure 4: A convection map formed by SuperDARN radars in the Northern Hemisphere, with coloured arrows representing the intensity and flow direction of plasma in the ionosphere. The swept region indicates the field of view of the Prince George Radar (PGR) ([Fear et al. 2017](#)).

SuperDARN detects radio wave reflections from magnetic field aligned plasma irregularities in the ionosphere. SuperDARN is a coherent scatter radar, therefore, if the irregularities are spaced with a distance of half the beam wavelength, constructive interference can occur and the wave is coherently scattered back towards the receiver. However, if the beam is reflected off the ground before returning to the radar this can produce a signal known as groundscatter.

The backscattered signal is received by SuperDARN and a pulsed sequence is returned. This sequence shows how ionospheric irregularities change in time and range. Range is divided into 45km segments known as range gates, along the length of each beam. The following parameters can be extracted from these pulsed sequences ([SuperDARN Tutorial](#)):

1. Power (dB)

The intensity of the radar beam reflected back to the receiver after being scattered off ionospheric irregularities.

2. Doppler Velocity ($m s^{-1}$)

The line of sight velocity of ionospheric plasma, measured along the beam direction. Positive Doppler velocity indicates motion away from the radar, and a negative velocity indicates motion towards the radar.

3. Spectral Width ($m s^{-1}$)

The measure of spread of the Doppler Velocity and a reflection of the persistence of a scattering region.

Figure 5 shows an example of these parameters plotted in range gates vs time.

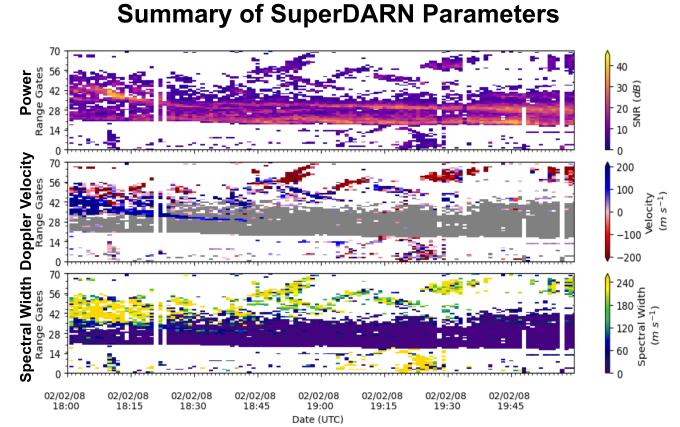


Figure 5: Example observations of the parameters extracted from measurements of plasma irregularities in the ionosphere by SuperDARN. Parameters include: power, Doppler velocity and spectral width. The grey flag in Doppler velocity represents groundscatter. Observations are from beam 5 of PGR, observed on 08/02/2008 between 18:00-19:45.

2.2.2 Coordinate Systems and AACGM

Whilst the radar receives pulsed sequences in range gates, viewing the data in Altitude-Adjusted Corrected Geomagnetic (AACGM) coordinates provides a fully consistent view across all 16 beams, ensuring that observations share a common axis of latitude ([Thayer School 2025](#)). This is because range gates are measured relative to the position of each antenna in the radar array. On the other hand, AACGM coordinates aim to provide an accurate representation of the Earth's non-dipolar magnetic field ([Shepherd 2014](#)).

2.2.3 Classifying Pulsed Ionospheric Flows (PIFs)

PMAFs can be classified in radar data from SuperDARN, where they are known as Pulsed Ionospheric Flows (PIFs). Studies have been conducted which have employed manual identification of PIFs by selecting specific observations with intense backscatter power or Doppler velocity increasing in magnetic latitude and time ([Provan and Yeoman 1999](#); [McWilliams et al. 2000](#); [Fear et al. 2017](#)). An example of this is shown in Figure 6. The poleward increase in ionospheric irregularities reflects the transport of magnetic flux across the polar cap, a direct consequence of magnetospheric convection.

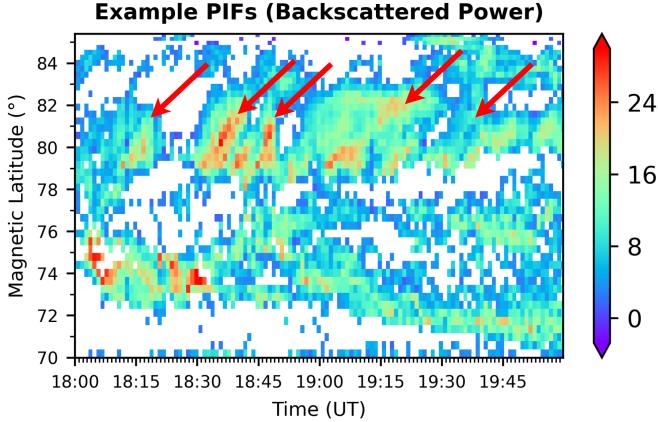


Figure 6: A magnetic latitude-time plot of backscattered power showcasing PIF events. Events are identified by the red arrows pointing to intense backscatter power increasing in magnetic latitude. The PIFs were observed by beam 10 of PGR on 08/02/2002 between 18:00 and 20:00 and first discovered by [Fear et al. \(2017\)](#).

2.3 SELF-SUPERVISED MACHINE LEARNING: SIMCLR

Machine learning techniques have shown significant promise in the classification of time-series data. Traditional Supervised Learning (SL) approaches require large, labelled data sets to train models. Unsupervised Learning (UL) aims to learn patterns/structures in data without the use of labels ([IBM 2024](#)). However, in the context of PIF classification, labelled data is limited, and previous attempts to use unsupervised learning have shown limitations in the speed and separation of high density clusters containing PIFs ([Robb et al. 2018](#)).

Self-Supervised Learning (SSL) is a subset of unsupervised machine learning which generates pseudo-labels from data. SSL methods differ from unsupervised learning in that they learn relative to a ground truth, similar to class labels in supervised learning ([Bergmann 2023](#)).

2.3.1 SimCLR Framework

An SSL approach that is particularly effective at classifying time-series data is SimCLR (Simple Framework for Contrastive Learning of Visual Representations) introduced by [Chen et al. \(2020\)](#). While SimCLR has not been applied to SuperDARN time-series data, it has found success in similar domains and can similarly be adapted to SuperDARN data. [Johnson et al. \(2021\)](#) used SimCLR to classify substorm auroral images and found a 10% boost in classification accuracy over benchmark methods. For general time-series data, pre-training using SimCLR has shown boosts in F1 scores and general classification benchmarks ([Zhang et al. 2022](#)).

SimCLR aims to learn similar representations of data by comparing augmented views (positive pairs) of the same datapoint, and maximising their similarity in their representation space, while pushing dissimilar samples away. In this way, the model is trained to identify samples which are similar without using explicit data labels.

After positive pairs are created, a neural network, typically a CNN extracts features from the images/time-series data. The CNN is the backbone of the SimCLR network and consists of:

- Fully connected layers, used in traditional neural networks. An image/time-series is input into the network and fully connected layers predict an output vector by learning weights associated with each layer as the model trains.
- Convolutional layers slide filters over input data to extract characteristic features from images. This results in an output feature/representation vector, also known as an embedding. Additional pooling layers help to down sample the input into lower dimensional fundamental features (IBM).

Combining these steps allows a CNN to extract characteristic features from data, subsequently placing features into a higher dimensional embedding space. In SimCLR, the feature vectors are projected into the contrastive loss function, which aims to minimise the distance between similar pairs of data in the embedding space ([Chen et al. 2020](#)).

An additional feature that can be applied to the CNN in a SimCLR network is a multi-head attention mechanism. Multi-head attention allows for machine learning models to focus on specific features of interest in a model, and has shown success in boosting performance in time-series classification ([Bergmann and Stryker 2024; Rafi and Woong-Ko 2021; El Zaar et al. 2025](#)). The method for multi-head attention is described in section 3.2.1.

3 METHOD

The following section outlines the methodology used to preprocess SuperDARN data, apply the SimCLR framework, cluster the embeddings and compute accuracy metrics. The pipeline is developed using python and ML models are developed using PyTorch. The method outlines a general framework, as such specific hyperparameter choices will be outlined in section 4 (Results).

3.1 PRE-PROCESSING

3.1.1 Extracting and Processing Records

SuperDARN data files are stored as compressed .bz2 files in 2 hour intervals. They must be decompressed and loaded using pyDARN, an open source python library used for data visualisation and input/output reading of SuperDARN files. Records are extracted from the loaded SuperDARN files and filtered by specific beam numbers. For each observation, the extracted records contain time (years/month/day/hour/minutes/second), range gates, backscatter power, Doppler velocity, spectral width and the radar station id. Time is converted to UTC datetime objects, and any missing/NaN measurements are filled with a padding value of -9999, significantly out of the range of typical large power/velocity/spectral width values ($\sim 45\text{dB}$, $\sim 1000\text{ms}^{-1}$, $\sim 1000\text{ms}^{-1}$). This process is iterated for all beams and records are stored in a data frame containing all observations.

3.1.2 Converting to Magnetic Latitudes

In order to maintain a consistent view of the data between different beams, as discussed in section 2.2.2, observation

coordinates must be converted. Range gates are mapped to geomagnetic latitudes using AACGM. The range gates of each beam intersect AACGM latitude and longitude lines, forming rectangular corners. Using the station ID, beam number and time of conversion, AACGM returns the coordinates of the corners, which can be averaged to obtain the magnetic latitude centre for each range gate. This ensures that the CNN in the SimCLR framework outlined in section 2.3.1 see consistent spatial structures.

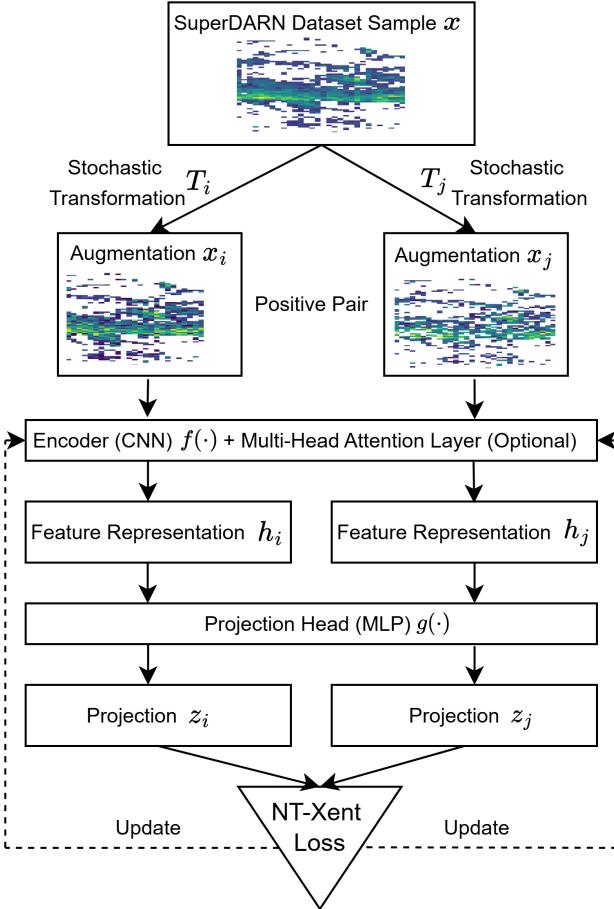


Figure 7: Schematic diagram of the SimCLR framework, illustrating stochastic data augmentations, the extraction of feature vectors using a base encoder with an optional MHA, projection head and contrastive loss optimisation for self-supervised learning.

3.1.3 Data Segmentation

PIF events occur over timescales of 10s of minutes (Provan and Yeoman 1999). In order to allow the SimCLR network to pick out PIF features, it is important that each sample fully contains a PIF, rather than multiple/fragmented events which may make it harder for the model to specialise to the characteristic poleward moving trajectory. As such, observations are segmented into 1 hour windows. To do this, valid data is selected conditioned upon them containing at least 100 data points. In order to ensure consistency across segments, each time step in the hour is resampled to 30 even time steps according to their nearest neighbour.

This results in an output tensor with dimensions [30 (time steps) x M (magnetic latitude bins) x 3 (power, velocity, spectral

width)]. The data is then further split into a train, test and validation set and stored into HDF5 files.

3.2 TRAINING

The following section outlines how the SimCLR network is trained using the pre-processed SuperDARN data. Although each observation provides 3 parameters (backscatter power, Doppler velocity and spectral width) PIFs are most visible in backscatter power observations. For simplicity, this study only trains on backscatter power values.

3.2.1 SimCLR Network

Training the SimCLR network occurs in multiple epochs where batches of data are fed to the network, allowing the base encoder to update its weights. A schematic of the SimCLR framework is illustrated in Figure 7. The training objective is to update the base encoder in the SimCLR network to correctly maximise the similarity between augmented pairs of the same sample. In this way, samples with similar spatio-temporal features (such as PIFs) should have high similarity scores and exist closer together in the embedding space. The framework used for SimCLR is outlined below and adapted from Chen et al. (2020).

a) **Data Augmentation:** Each sample x is augmented to produce two correlated views, through stochastic transformations T_i and T_j , creating a "positive pair" of samples x_i and x_j .

Data is stochastically transformed using six augmentation methods outlined below. Each transformation is applied with some probability, defined by respective hyperparameters. A diverse range of augmentations is critical to improve the robustness of the SimCLR network and allow the model to generalise to a wider variety of features (Chen et al. 2020). The following augmentations have been chosen because their combination results led to stable model training and prevented severe divergence during training:

ai) *Random Noise*: Random noise is added to all data points from a random normal distribution with mean 0 and standard deviation $\sigma = \text{Noise Strength} \cdot \text{Range}$. The noise strength is a tuneable hyperparameter and the range $= x^{\max} - x^{\min}$.

aii) *Scaling*: Data points are multiplied by random scaling values pulled from a random uniform distribution with lower and upper bounds defined by a tuneable hyperparameter called the scale range.

aiii) *Saturation*: This augmentation aims to emulate traditional image saturation. Data is scaled according to the following transformation: $x_{i,j} = \mu_x + S \cdot (x - \mu_x)$. Where μ_x is the average backscatter power value for a sample and S is a hyperparameter defined as the saturation factor. The aim of this transformation is to apply more scaling to more extreme data values.

avi) *Masking*: Valid data points are randomly masked (set to the negative padding value) with a defined probability.

av) *Swapping*: If two adjacent data points are valid they may be swapped according to a random probability.

avi) *Translation*: Data can also be translated in magnetic latitude or time. A random shift is chosen up to a parameter called max

shift. For the y direction, max shift = $2 \cdot$ max shift in the x direction to account for the larger range of the y axis. Data can be shifted up to 3 times, until a maximum number of points are removed.

A visualisation of the different augmentations is displayed in Figure 8.

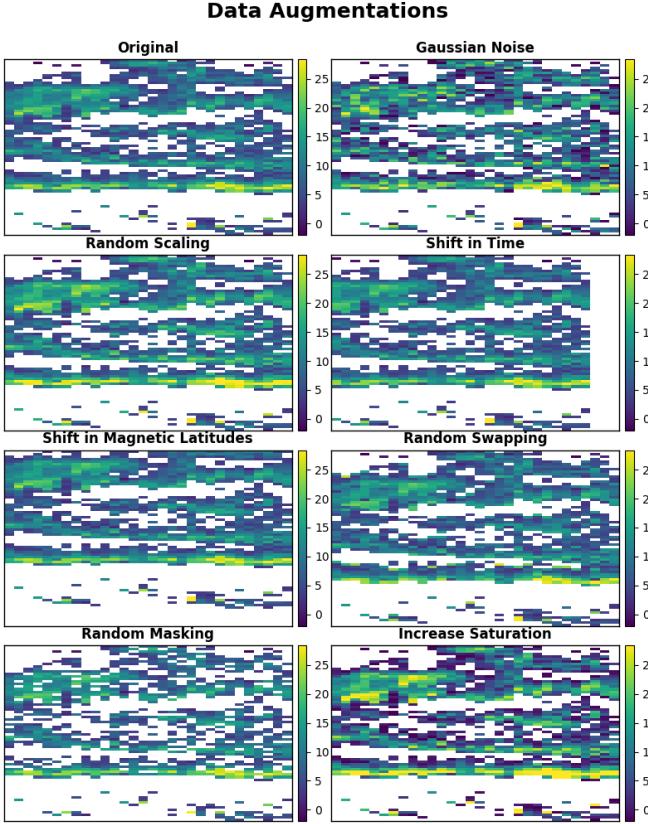


Figure 8: Examples of the applied augmentations to samples from the SuperDARN dataset. Augmentation strengths are intentionally amplified in this figure and not accurate to the specific hyperparameters used during training.

In each epoch, a batch of data with augmentations is assembled using a contrastive collate function. This function assembles a batch by stacking the augmented views into a single tensor with dimensions $[2 \times \text{batch_size}, 30, M]$. In this step, augmented data is masked so that missing samples are replaced with a negative padding value, and then data is normalised. Data is normalised relative to the global mean (μ) and standard deviation (σ) of the data and is outlined in equation 1.

$$\mathbf{x}_{i,j}^{\text{norm}} = \frac{\mathbf{x}_{i,j} - \mu}{\sigma} \quad (1)$$

The collated batch is then passed through the following stages of the network.

b) **Base Encoder:** The base encoder is a Convolutional Neural Network (CNN) $f(\cdot)$ responsible for extracting the following representation vectors from the augmented samples: $\mathbf{h}_i = f(\mathbf{x}_i)$, $\mathbf{h}_j = f(\mathbf{x}_j)$. The encoder outputs a feature map with 512 dimensions

bi) **Multi-Head Attention:** Optionally, the network can also be run with a multi-head self-attention layer, following the

architecture described by (Vaswani et al. 2017). Multi-head attention runs several attention layers in parallel. Each attention layer aims to determine the relative significance of different positions (tokens) in a sequence, and can subsequently be applied to sequences of features extracted by the CNN. In this mode, the feature map extracted by the CNN is flattened into a sequence of tokens.

This study uses self attention, therefore each token is projected into a query (Q), key (K) and value (V) vector. The query aims to look for a specific pattern, the key states if a token has the specific pattern and the value contains the actual information in the token. More weight is applied to keys aligned with the query and less relevant queries are given a weighting closer to 0. A boolean mask is also input showing the positions of any padding (invalid) values. In this context, the goal of attention is for the base encoder to focus more on the valid values of backscatter power and understand their spatial and temporal relationships. The tokens and mask are fed into the single multi-head self-attention layer with 4 heads.

The attention tokens are pooled (averaged) and produce a final 512 dimension feature map. This vector is passed into the projection head.

c) **Projection Head:** Representation vectors are then passed through the projection head, in this case, a Multi Layer Perceptron (MLP) $g(\cdot)$ is chosen. This results in the projections: $\mathbf{z}_i = g(\mathbf{h}_i)$, $\mathbf{z}_j = g(\mathbf{h}_j)$. An MLP is a traditional neural network, in that it consists of multiple fully connected layers. The goal of this step is to project the representation vectors into a lower dimensional space (128), suitable for the NT-Xent contrastive loss function. After projection, L2 normalisation is applied, ensuring all projected vectors have unit length 1.

d) **NT-Xent Contrastive Loss Function:** Finally, the projected representations are input into the Normalised Temperature-Scaled Cross-Entropy (NT-Xent) Loss Function. This function aims to carry out contrastive prediction, where for a set of observation pairs, the function identifies positive pairs and aims to maximise their similarity, whilst pushing apart dissimilar samples. The NT-Xent loss function is defined in equation 2.

In order to quantify the similarity between projected representations a similarity matrix is formed using cosine similarity. Cosine similarity is defined as $\text{sim}(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i^T \cdot \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|}$. The objective during training is to maximise $\text{sim}(\mathbf{z}_i, \mathbf{z}_j)$ for positive pairs. The parameter τ defines temperature, and is used to adjust the sensitivity of the similarity scores by scaling cosine similarity. Temperature is defined in the range $0 \leq \tau < 1$, as such a higher temperature results in less contrast and a lower temperature results in increased contrast between similar pairs.

$$\mathcal{L}(\mathbf{x}_i, \mathbf{x}_j) = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)} \quad (2)$$

The training loop uses the Adam optimiser to update the weights in the CNN encoder and MLP projection head. This is achieved by backpropagating the contrastive loss computed between pairs of samples, so that it is minimised for a set of optimal weights. The speed of convergence can be adjusted using the learning rate. To prevent the model overfitting to the training data, early

stopping is also employed. To do this, a value called patience is defined. If the validation loss does not improve for a number of epochs defined by patience, the model stops training.

By training with this framework SimCLR is able to bring the representation vectors of similar samples closer together in the learned embedding space whilst pushing apart dissimilar samples.

3.3 ACCURACY METRICS

The following section outlines accuracy metrics used to understand the effectiveness of SimCLR in identifying PIFs.

3.3.1 Top-K Accuracy

Top-K accuracy helps to show the effectiveness of the model in separating similar or dissimilar data points in the embedding space. For an anchor sample x_i , top-K accuracy is computed by calculating the percentage of times the correct pair x_j is identified in the top k embeddings along the row of the similarity matrix. During training, Top-K accuracy is computed at the end of each epoch using the validation set. High top-k accuracies indicate that the model is correctly structuring the embedding space.

3.3.2 Visualisation

In order to ensure that the model is correctly maximising the similarity between PIF observations, visualising the top-k closest samples in the embedding space is a useful metric.

This can be done by loading the pre-processed test set into the trained base encoder without applying augmentations to extract their embeddings. Finally, an anchor sample is chosen in the embedding space, and the top-k closest embeddings are plotted.

3.3.3 Clustering

Clustering can also be applied to further probe the structure of the learned representation space. In this study, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is applied as it specialises in grouping data by density. DBSCAN was first introduced by Ester et al. (1996). DBSCAN measures the density around points using MinPts which specifies the minimum number of points needed to form a cluster and Epsilon which defines the radius of the neighbourhood around a certain point. Embeddings are extracted from the trained base encoder using the test set and DBSCAN is applied. In order to understand the quality of each cluster, the silhouette score can be calculated using equation 3

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Where $a(i)$ is the average distance between data points i in the same cluster and $b(i)$ is the average distance to the nearest external cluster. The silhouette score has bounds $-1 \leq S(i) < 1$. A value closer to 1 indicates a coherent cluster. $S(i) \approx 0$ indicate

overlapping clusters, and $S(i) \approx -1$ indicates incorrectly clustered samples ([Scikit-learn](#)).

In order to visualise the clusters, a Principal Component Analysis (PCA) can be computed. PCA is a dimensionality reduction technique to reduce the embeddings to a 2D space. PCA identifies the top 2 with maximal variance known as PC1 and PC2. A scatter plot of PC2 vs PC1 can display the locations and sizes of clustered embeddings ([Maćkiewicz and Ratajczak 1993](#)).

4 RESULTS

This section presents the results from training the SimCLR network on SuperDARN data, including performance metrics and DBSCAN clustering. Performance is compared between 2 experiments, running the model with a multi-head attention layer in the base encoder and without. All values are quoted to 3 s.f. unless otherwise specified.

4.1 TRAINING PERFORMANCE

The data used to train the SimCLR network was observed by the SuperDARN Hankasalmi radar in Finland between 19/06/1995 00:00 - 1996/01/00 00:00. This period contained 2245 2 hour observations. A train/test/validation split of 70%/15%/15% was chosen. After pre-processing this resulted in a total of 62,288 1 hour segments split into 43,601(train)/9343(validation)/9344(test). Data from this period was specifically chosen due to its overlap with a study completed by [Provan and Yeoman \(1999\)](#), as this study produced a sample of labelled PIFs, which are used for analysis in section 4.3. Training was completed using a g5.xlarge instance on Amazon Web Services (AWS). The model without attention trained for 74 epochs over 2 hours and 28 minutes whereas the model with attention trained for 86 epochs over 2 hours and 57 minutes. The varying lengths in epochs are due to a chosen patience value of 5, which resulted in early stopping to stop overfitting. The following parameters remained consistent across both experiments:

A standard learning rate of 1×10^{-3} is chosen to ensure stability while training. [Chen et al. \(2019\)](#) found that a larger batch size helps provide SimCLR with more negative pairs, resulting in improved model generalisation. Subsequently, a large batch size of 256 is chosen and a temperature (τ) of 0.5 is chosen.

The chosen augmentation hyperparameters are displayed in table 1. All parameters had a 50% probability of application. These hyperparameters were chosen as they resulted in a stable training run. Increasing certain hyperparameters such as noise resulted in a model which did not converge.

Augmentation	Strength / Range
Gaussian noise	0.40
Scaling	0.9–1.1
Saturation	1.3
Masking	0.10
Swapping	0.10
Translation	Max shift = 4; removed 400 pts

Table 1: Augmentation hyperparameters used during training, chosen for stable model convergence.

4.1.1 Loss and Top-k Accuracy

The top-k accuracy and loss over all trained epochs are displayed in figure 9. The training and validation loss curves exhibit a stable and consistent decrease over all epochs indicating a smooth convergence across both experiments.

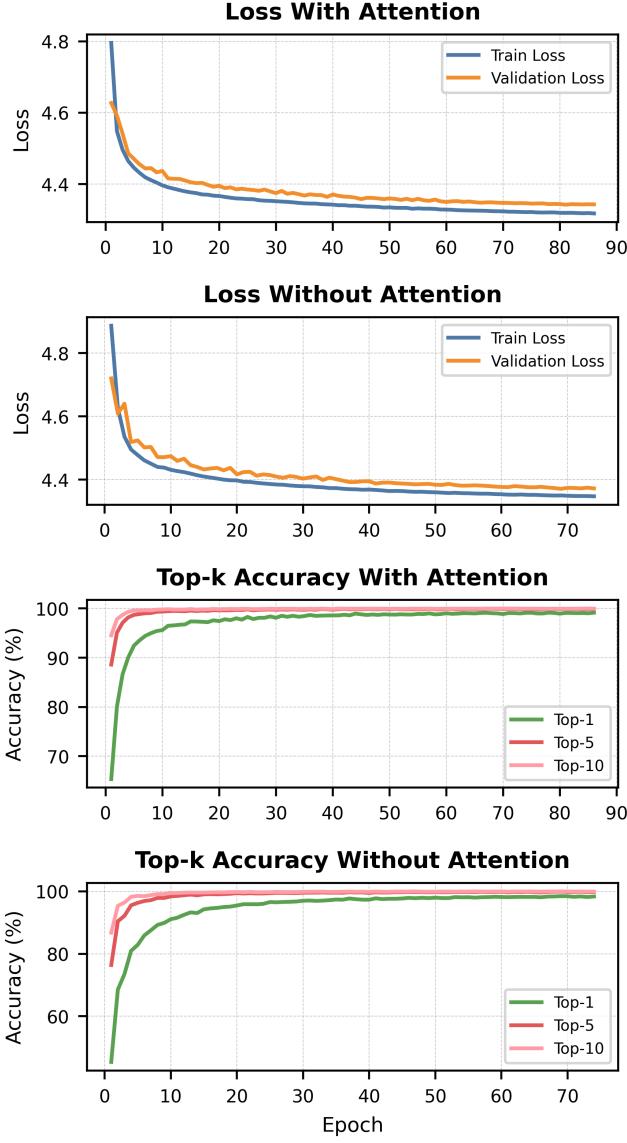


Figure 9: Training and validation loss (top) and top-k accuracy (bottom) vs epochs for the non-attention model (right) and the attention model (left).

Top-k accuracy increases sharply to > 95% after epoch 5, and then maintains a stable increase to a high accuracy. Notably, the final top-1 accuracy without attention was 98.3% and with attention was 99.15%. The initial top-1 accuracy was 19.94% higher using attention mechanisms and plateaued much earlier at epoch 40 compared to epoch 50 without attention mechanisms.

4.2 DBSCAN CLUSTERING

DBSCAN clustering was applied with epsilon = 0.4 and minimum points = 20 for both experiments on embeddings from the training set with 9344 samples. These parameters were chosen as they provided the best separated and coherent clusters without

excessive fragmentation. PCA diagrams for both experiments are displayed in figure 10.

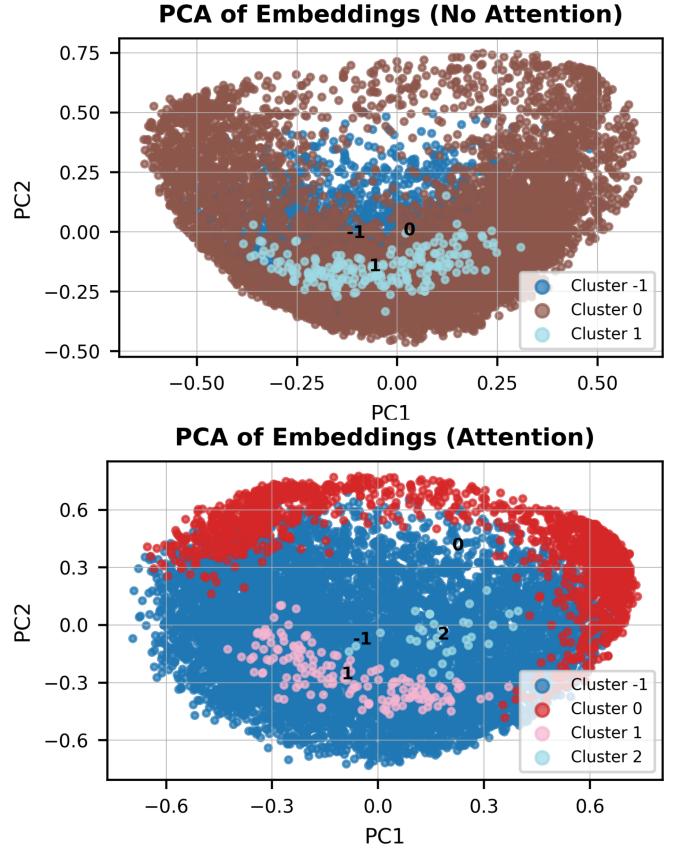


Figure 10: PCA of DBSCAN clustered 512D embeddings extracted from the test data set using a trained base encoder without attention (top) and with attention (bottom). The individual clusters are highlighted across both experiments after running DBSCAN.

3 clusters were identified without attention and 4 clusters were identified with attention. The number of points in each cluster and their silhouette scores are outlined in table 2. Cluster -1 and 0 were the largest between both experiments. After visual inspection of the backscatter power, both clusters consisted of observations with no coherent relationship suggesting they are noise. Cluster 2 also consisted of incoherent observations.

Cluster	Attention		No Attention	
	Count	S(i)	Count	S(i)
-1	7624		2017	
0	1535	0.236	7128	0.0043
1	152	0.529	199	0.527
2	33	0.581		

Table 2: The number of points in the clusters highlighted by DBSCAN across both experiments, and their respective silhouette scores. Clusters 0, 1, and 2 consist mostly of incoherent observations, however, cluster 1 contains some observations with PIF characteristics.

However, after inspecting cluster 1, some observations indicated the characteristics of PIFs. The following 3 types of observations were resolved:

1. **PIFs:** These samples exhibited the qualities of PIFs. Observations had poleward moving forms of intense backscatter power. However, in order to classify these as events, the

Doppler velocity and spectral width were referenced. This is because intense poleward moving power may also be a result of groundscatter. As such, to confirm an observation as a PIF, high Doppler velocity and broad spectral width must also be observed.

2. Groundscatter: Samples had increases in poleward moving forms of intense backscatter, however they had narrow spectral width and low Doppler velocity. Figure 11 shows an example of a sample considered an event vs groundscatter. The grey flag in Doppler velocity indicates groundscatter.

3. Non-Events: Samples which did not exhibit poleward moving forms of intense backscatter/Doppler velocity/spectral width are flagged as non-events.

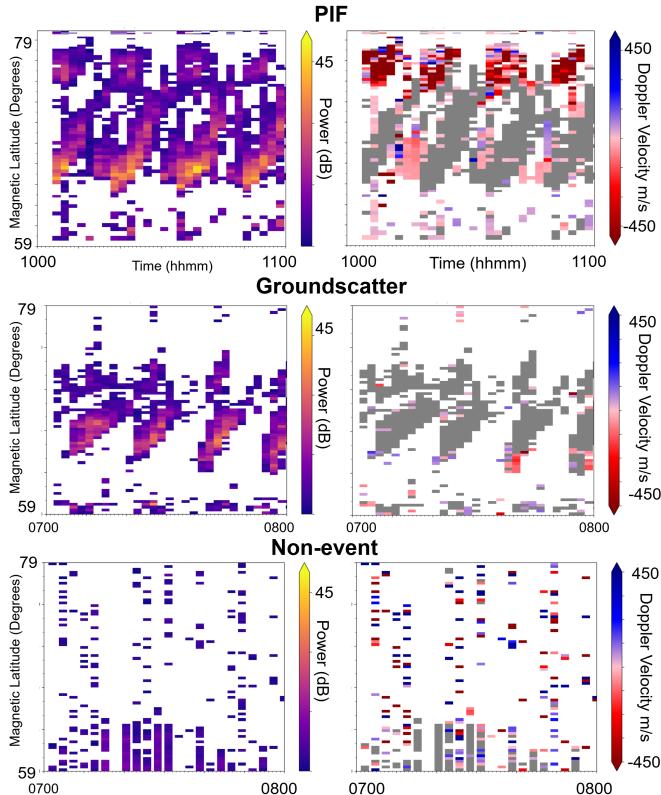


Figure 11: 3 Examples of visually classified samples from cluster 1 indicating a PIF event with high poleward moving Doppler Velocity (top) groundscatter (middle) and a non-event (bottom). Both types of observation are viewed in power (left) and Doppler velocity (right) to ensure an accurate classification.

Figure 11 shows a comparison of a PIF, groundscatter and non-event. The PIF sample exhibits short burst (< 10 minutes) of high intensity poleward moving backscatter, with high Doppler velocities ($> 500 \text{ ms}^{-1}$), consistent with those described in section 2.2.3. The accuracies for PIFs, groundscatter and non-events in cluster 1 are displayed in figure 12.

Across both experiments, most observations in cluster 1 are groundscatter $\approx 52\%$, followed by PIFs $\approx 30\%$ and non-events $\approx 18\%$. The non-attention model has a cluster size of 199 whereas the attention model has a cluster size of 152. Notably, there is a 3.2% increase in the detection of PIFs in the cluster using attention mechanisms. Cluster 1 also had the highest silhouette score without attention and the second highest with attention. The score for cluster 1 was also similar (≈ 0.53

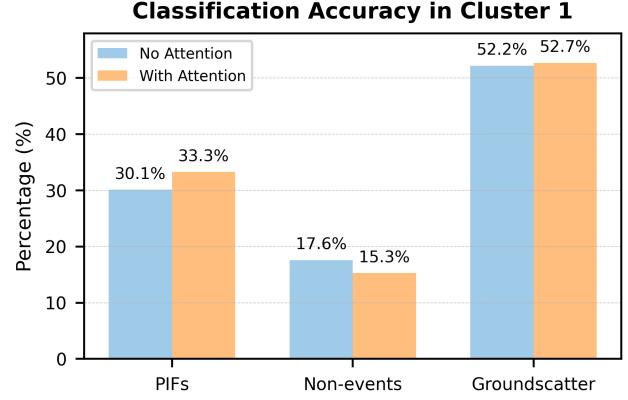


Figure 12: The classification accuracy in cluster 1 for PIF, groundscatter and non-events across 2 training runs with an attention model (purple) and without (blue). Using attention mechanisms introduces a slight increase in PIF classification accuracy (+3.2%)

(2s.f.)) for both experiments. Cluster 1 had a similar location and form in the 2D space across both experiments.

4.3 EXAMPLE EVENT

To further probe the structure and consistency of the embedding space an anchor sample known to be a PIF is chosen and its top 50 neighbours are inspected. This anchor sample took place on 24/09/1995 and was first identified by [Provan and Yeoman \(1999\)](#). The anchor sample is shown in figure 13.

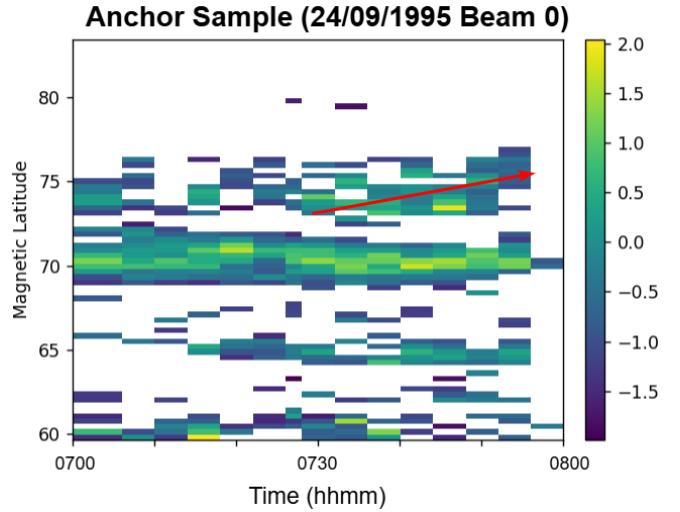


Figure 13: A sample identified as a PIF by [Provan and Yeoman \(1999\)](#) indicating a slower event relative to those classified in cluster 1, the specific flow is indicated along the arrow. This sample is used as an anchor to probe the structure of the embedding space by classifying its top-50 neighbours. The colour bar illustrates normalised power values, subsequently it has a range between -2 and 2.

This sample was chosen as it takes a different form to the PIFs identified in cluster 1. The PIF moves poleward on a longer time scale spanning approximately 10° in 30 minutes relative to samples in cluster 1 which spanned 10° in 10 minutes. This provides insights into the effectiveness of SimCLR in maximising similarity between a wider range of PIF patterns

beyond those identified in cluster 1. The location of the anchor in the PCA relative to cluster 1 is displayed in figure 14, along with the top 50 neighbours around the anchor.

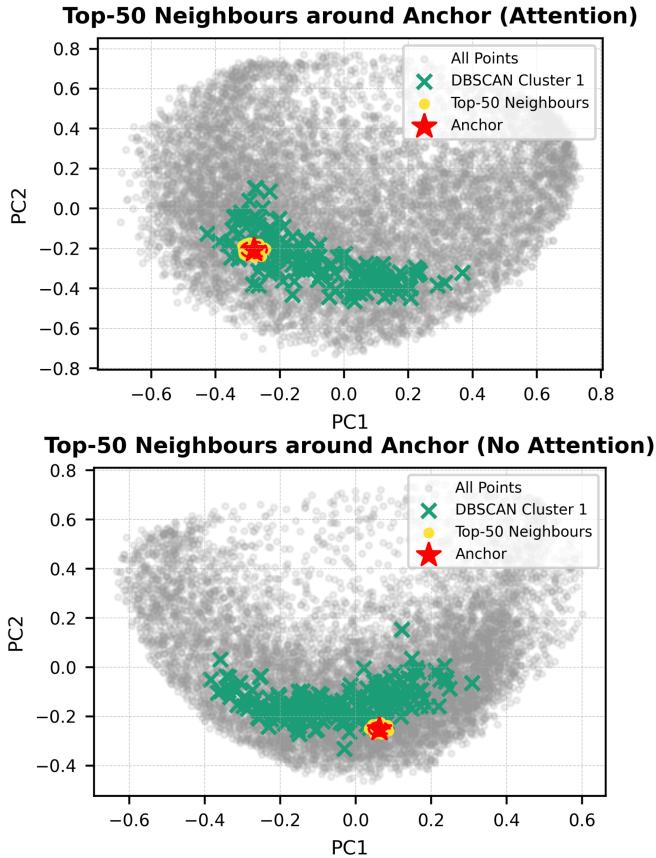


Figure 14: A PCA indicating the top-50 nearest neighbours (orange points) around the chosen anchor PIF (red star). The points of cluster 1 are also highlighted for the attention model (bottom) and non-attention model (top).

By visually inspecting the 50 neighbours around the anchor, 8 were identified as PIF events using the non attention model and 6 events using the attention model. Among the identified PIFs, with attention 3 samples were also in cluster 1 and without attention 5 samples were also in cluster 1. The PIF samples found in the top-50 neighbours exhibited similar spatio-temporal patterns as the anchor sample. On average the similar PIFs spanned approximately $10^\circ - 15^\circ$ in a time range of 25-35 minutes.

5 DISCUSSION

This section provides interpretations of the results and an evaluation of the efficiency of SimCLR in identifying PIFs with and without attention mechanisms.

It is evident that the top-k accuracies increase to large values across both experiments. Whilst there is no classical overfitting between the training and validation loss, the high top-k accuracies indicate that the model has overfit to the positive pairs created through data augmentation, subsequently reducing its generalisation to unseen data. Whilst clustering may provide a different view on the learned representations of the model, the unrealistically high top-k accuracy is likely an indicator that the augmentation strengths may need increasing, or alternate

augmentations may be required to improve the robustness of the model.

The attention model had higher initial (+19.94%) and final (+0.85%) top-1 accuracies than the non-attention model. Although these increases were small, it indicates that using attention may help the model to bring semantically similar observations closer together in the embedding space by focusing more on the spatio-temporal patterns in the data. However, the trade-off in attention mechanisms is that they took more epochs (+12) and time (+28 minutes) to converge to a stable model which shows limitations in computational efficiency. Only considering accuracy, the gains using attention mechanisms may not be worth the additional computational power.

While there are signs that the model overfit to positive pairs, the ability of the model to group similar observations and capture meaningful representations of the data is evident by the results of DBSCAN clustering. Cluster 1, separated by DBSCAN in both experiments contained a group of observations with poleward moving forms of intense backscattered power with a duration of approximately 10 minutes, occurring approximately every 15 minutes. This is consistent with the description of reconnection timescales by [Russell and Elphic \(1978\)](#), where reconnection events are described to occur over a matter of minutes.

This cluster had a moderate silhouette score (≈ 0.53 (2.s.f.)) across both experiments, indicating that it is not fully separated from other clusters and may be moderately cohesive. This is also evident by the fact that the PCA space exhibits obvious overlapping between clusters. This is likely a result of overfitting to positive pairs, as the model may fail to push apart negative samples in the embedding space, resulting in an unstructured embedding space. A limitation of previous studies such as [Robb et al. 2018](#) was the separation of clusters and this still remains a challenge in this study. However, this limitation is not necessarily inherent to SimCLR but rather a result of suboptimal augmentation strategies.

However, within the cluster approximately 30% of the samples exhibited the characteristics of PIFs across both experiments. As the cluster was moderately cohesive, most samples showed patterns in backscattered power which were predominantly poleward moving, however, the majority of the poleward moving samples were identified to be groundscatter. Even though the majority of the cluster was groundscatter, the patterns in backscatter power were predominantly poleward moving. In this capacity, SimCLR has been effective in bringing together observations with poleward moving power, however, without access to alternate parameters such as Doppler velocity and spectral width the model is unable to exclusively bring poleward moving forms with high velocity or broad spectral width together, which are significant characteristics of PIFs. This limitation could be solved by using multi-modal learning. Multi-modal learning explores the interactions between different data types, by generating combined representations of data through a process known as fusion ([Kline et al. 2022](#)). Applied to SimCLR, multi-modal learning has shown success in benefiting downstream generalisation ([Zhang et al. 2023](#)) and reducing the reliance on augmentations to improve model performance ([Jain et al. 2022](#)). As such, applying multi-modal contrastive learning may help to exclude erroneous/groundscatter data from

the embedding space associated with poleward moving auroral forms.

Using attention mechanisms, cluster 1 had 47 less data points compared to the non-attention model. There was also a small increase (+3.2%) in the classification accuracy of PIFs using attention mechanisms. The clusters were classified by visually inspecting the samples, as such, the increase in PIF classification accuracy may be a result of a subjective variability due to inspection. The smaller sample size in cluster 1 when using attention also increases the potential for variability in the accuracy, as such a slight improvement in PIF classification is likely not a significant trend. However, the fact that cluster 1 in the attention model was smaller indicates that attention mechanisms may have learned overly specific features resulting in a smaller cluster. This suggests that while attention may help to bring semantically similar samples together as evidenced by higher top-k accuracies, it may over-specialise to specific types of spatio-temporal patterns in the samples.

To further investigate this, the quality of the embedding space was probed using the PIF identified by Provan and Yeoman (1999). This sample had a different spatio-temporal pattern to those observed in cluster 1. In the top-50 nearest embeddings around this sample, with attention, 12% were identified as PIFs, without attention, 16% were identified as PIFs. These values are significantly lower than the classification accuracy values for PIFs in cluster 1. As such, for PIFs with less dramatic spatio-temporal trends such as on a longer time scale, SimCLR has some difficulty maximising the similarity between these samples. Although there was some overlap with the samples in cluster 1, it was not sufficient to conclude that SimCLR reliably captures all poleward-moving forms, particularly those with more gradual or extended dynamics.

To address this specific limitation, a strategy known as fine-tuning can be implemented. Fine-tuning takes a pre-trained encoder, and adapts it using labels from a clustered data by training with a supervised loss function (Liu and Alavi 2024). This would allow for the model to further specialise to the classification of PIF events. This strategy was tried by project collaborator Smith (2025), where test set accuracies in the primary cluster containing PIFs were found to be 42.4% PIFs, 46.6% groundscatter and 11% non-events. This shows a considerable improvement over the classification accuracy of PIFs using attention in this study of 33.3%. However, a large amount of groundscatter was still classified, as such, future studies may aim to incorporate labelled PIFs with a wider variety of spatio-temporal patterns.

6 CONCLUSION

This study applied self-supervised contrastive learning to automate the identification of Pulsed Ionospheric Flows (PIFs) from SuperDARN radar data. By identifying a higher volume of PIFs, classifications could be used help to answer questions about the dynamics of magnetospheric convection. The study trained a SimCLR network using unlabelled observations from the Hankasalmi SuperDARN radar from 1995 whilst testing the effectiveness of multi-head attention on the performance of the trained model.

This study found that SimCLR converges stably to high top-1 accuracies of 98-99% using attention and non-attention models. Upon applying DBSCAN clustering a moderately coherent cluster emerged with an average silhouette score (≈ 0.53) in which $\sim 30\%$ of samples were visually identified as PIF events, $\sim 52\%$ as groundscatter and $\sim 18\%$ as non-events. Although groundscatter dominates, the model successfully grouped poleward moving signatures of bursty magnetic reconnection events. Using a single multi-head attention layer improved early convergence and raised the final top-1 score by $\sim 0.9\%$, but required +12 epochs to train (+ 28 minutes) and resulted in a smaller cluster containing PIFs, with only a 3.2% increase in PIF classification accuracy.

The high top-k accuracies and low cluster silhouette scores indicate the model overfit to augmented positive pairs. Subsequently, future work should aim to diversify the strengths/types of augmentations used. Whilst downstream DBSCAN clustering did identify some PIFs, the high classification of groundscatter indicates the model would benefit from multi-modal training by including Doppler velocity and spectral width parameters and not just backscatter power. The model lacked generalisation to PIFs with varying temporal patterns, as for a slower (~ 30 minutes) labelled PIF, only 14% (across both models) of the 50 nearest neighbours were also PIF events. Studies such as Smith (2025) have shown that fine-tuning using labelled cluster data increased PIF classification accuracy (+12%). Future work could aim to include labelled data, with a varied range of spatio-temporal patterns to improve model generalisation. By addressing these limitations, the SimCLR framework can be further adapted to produce an efficient and fully automated approach to identifying PIF signatures.

REFERENCES

- Axford, W.I. (1969). Magnetospheric convection. *Reviews of Geophysics*, 7(1, 2), pp.421–421. doi:<https://doi.org/10.1029/rg007i001p00421>.
- Bergmann (2023). What Is Self-Supervised Learning? [online] Ibm.com. Available at: <https://www.ibm.com/think/topics/self-supervised-learning>.
- Bergmann and Stryker (2024). Attention mechanism. [online] Ibm.com. Available at: <https://www.ibm.com/think/topics/attention-mechanism>.
- Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. arXiv:2002.05709 [cs, stat]. [online] Available at: <https://arxiv.org/abs/2002.05709>.
- Chisham, G., Lester, M., Milan, S.E., Freeman, M.P., Bristow, W.A., Grocott, A., McWilliams, K.A., Ruohoniemi, J.M., Yeoman, T.K., Dyson, P.L., Greenwald, R.A., Kikuchi, T., Pinnock, M., Rash, J.P.S., Sato, N., Sofko, G.J., Villain, J.-P. . and Walker, A.D.M. (2007). A decade of the Super Dual Auroral Radar Network (SuperDARN): scientific achievements, new techniques and future directions. *Surveys in Geophysics*, 28(1), pp.33–109. doi:<https://doi.org/10.1007/s10712-007-9017-8>.
- Dai, L., Zhu, M., Ren, Y., Gonzalez, W., Wang, C., Sibeck, D., Samsonov, A., Escoubet, P., Tang, B., Zhang, J. and Branduardi-Raymont, G. (2024). Global-scale magnetosphere convection driven by dayside magnetic reconnection. *Nature Communications*, [online] 15(1), p.639. doi:<https://doi.org/10.1038/s41467-024-44992-y>.
- Dungey, J.W. (1961). Interplanetary Magnetic Field and the Auroral Zones. *Physical Review Letters*, 6(2), pp.47–48. doi:<https://doi.org/10.1103/physrevlett.6.47>.
- El Zaar, A., Mansouri, A., Benaya, N., Bakir, T. and El Allati, A. (2025). Hybrid Transformer-CNN architecture for multivariate time series forecasting: Integrating attention mechanisms with convolutional feature extraction. *Journal of Intelligent Information Systems*. doi:<https://doi.org/10.1007/s10844-025-00937-5>.
- Ester, M., Xu, X., Sander, J. and Kriegel, H.-P. (2025). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Second International Conference on Knowledge Discovery and Data Mining (KDD'96). Proceedings of a conference held August 2-4, [online] pp.226–331. Available at: <https://ui.adsabs.harvard.edu/abs/1996kddm.conf.226E/abstract> [Accessed 26 Apr. 2025].
- Fear, R.C., Trenchi, L., Coxon, J.C. and Milan, S.E. (2017). How Much Flux Does a Flux Transfer Event Transfer? *Journal of Geophysical Research: Space Physics*, 122(12), pp.12, 310–12, 327. doi:<https://doi.org/10.1002/2017ja024730>.
- Greenwald, R.A., Baker, K.B., Dudeney, J.R., Pinnock, M., Jones, T.B., Thomas, E.C., Villain, J.-P. ., Cerisier, J.-C. ., Senior, C., Hanuse, C., Hunsucker, R.D., Sofko, G., Koehler, J., Nielsen, E., Pellinen, R., Walker, A.D.M., Sato, N. and Yamagishi, H. (1995). DARN/SuperDARN. *Space Science Reviews*, 71(1-4), pp.761–796. doi:<https://doi.org/10.1007/bf00751350>.
- IBM (2024). What Is Machine learning? [online] Ibm.com. Available at: <https://www.ibm.com/think/topics/machine-learning>.
- Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33. doi:<https://doi.org/10.1007/s10618-019-00619-1>.
- Jain, U., Wilson, A., Gulshan, V. and Research, G. (2022). Multimodal contrastive learning for remote sensing tasks. [online] Available at: https://sslneurips22.github.io/paper_pdfs/paper_34.pdf [Accessed 26 Apr. 2025].
- Johnson, J.W., Hari, S., Hampton, D., Connor, H.K. and Keesee, A. (2021). A Contrastive Learning Approach to Auroral Identification and Classification. 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), [online] pp.772–777. doi:<https://doi.org/10.1109/icmla52953.2021.00128>.
- Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F. and Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, [online] 5(1), pp.1–14. doi:<https://doi.org/10.1038/s41746-022-00712-8>.
- Kunduri, B.S.R., Baker, J.B.H., Ruohoniemi, J.M., Thomas, E.G. and Shepherd, S.G. (2022). An Examination of SuperDARN Backscatter Modes Using Machine Learning Guided by Ray-Tracing. *Space Weather*, 20(9). doi:<https://doi.org/10.1029/2022sw003130>.
- Liu, E., Hu, H., Liu, J., Teng, X. and Qiao, L. (2019). Predicting SuperDARN cross polar cap potential by applying regression analysis and machine learning. *Journal of Atmospheric and Solar-Terrestrial Physics*, 193, p.105057. doi:<https://doi.org/10.1016/j.jastp.2019.105057>.
- Lockwood, M., Per Even Sandholt, Stanley and Takasi Oguti (1989). Interplanetary magnetic field control of dayside auroral activity and the transfer of momentum across the dayside magnetopause. 37(11), pp.1347–1365. doi:[https://doi.org/10.1016/0032-0633\(89\)90106-2](https://doi.org/10.1016/0032-0633(89)90106-2).

- Maćkiewicz, A. and Ratajczak, W. (1993). Principal components analysis (PCA). *Computers and Geosciences*, 19(3), pp.303–342. doi:[https://doi.org/10.1016/0098-3004\(93\)90090-r](https://doi.org/10.1016/0098-3004(93)90090-r).
- McWilliams, K.A., Yeoman, T.K. and Provan, G. (2000). A statistical survey of dayside pulsed ionospheric flows as seen by the CUTLASS Finland HF radar. *Annales Geophysicae*, 18(4), pp.445–453. doi:<https://doi.org/10.1007/s00585-000-0445-8>.
- Sandholz, Deehr, C., Egeland, A., Bjørn Lybekk, Viereck, R. and Romick, G.J. (1986). Signatures in the dayside aurora of plasma transfer from the magnetosheath. 91(A9), pp.10063–10063. doi:<https://doi.org/10.1029/ja091ia09p10063>.
- Provan, G. and Yeoman, T.K. (1999). Statistical observations of the MLT, latitude and size of pulsed ionospheric flows with the CUTLASS Finland radar. *Annales Geophysicae*, 17(7), pp.855–867. doi:<https://doi.org/10.1007/s00585-999-0855-1>.
- Rafi, T.H. and Woong-Ko, Y. (2022). HeartNet: Self Multi-Head Attention Mechanism via Convolutional Network with Adversarial Data Synthesis for ECG-based Arrhythmia Classification. *IEEE Access*, pp.1–1. doi:<https://doi.org/10.1109/access.2022.3206431>.
- Ribeiro, A.J., Ruohoniemi, J.M., Baker, J.B.H., Clausen, L.B.N., de Larquier, S. and Greenwald, R.A. (2011). A new approach for identifying ionospheric backscatter in midlatitude SuperDARN HF radar observations. *Radio Science*, 46(4), p.n/a-n/a. doi:<https://doi.org/10.1029/2011rs004676>.
- Rijnbeek, R.P., Cowley, H., Southwood, D.J. and Russell, C.T. (1982). Observations of reverse polarity flux transfer events at the Earth's dayside magnetopause. *Nature*, [online] 300(5887), pp.23–26. doi:<https://doi.org/10.1038/300023a0>.
- Robb, E. (2018). Space@VT GSOC - Esther Robb. [online] Github.io. Available at: https://vtsuperdarn.github.io/clustering_superdarn_data/. Russell, C.T. and Elphic, R.C. (1978). Initial ISEE magnetometer results: magnetopause observations. *Space Science Reviews*, 22(6). doi:<https://doi.org/10.1007/bf00212619>.
- Russell, C.T., Luhmann, J.G. and Strangeway, R.J. (2016). Space Physics. Cambridge University Press. Available at:<https://ui.adsabs.harvard.edu/abs/2016spai.book.....R/abstract>
- scikit-learn (2019). Selecting the number of clusters with silhouette analysis on KMeans clustering — scikit-learn 0.21.2 documentation. [online] Scikit-learn.org. Available at: https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html.
- Shepherd, S. (2014). Altitude-adjusted corrected geomagnetic coordinates: Definition and functional approximations. 119(9), pp.7501–7521. doi:<https://doi.org/10.1002/2014ja020264>.
- SuperDARN (2021). About :: Tutorials Menu. [online] Superdarn.ca. Available at: <https://superdarn.ca/tutorials> [Accessed 26 Apr. 2025].
- SuperDARN (2021). The Super Dual Auroral Radar Network (SuperDARN). [online] Available at: <https://superdarn.ca/> [Accessed 26 Apr. 2025].
- Thayer School (2025). AACGM. [online] Dartmouth.edu. Available at: <https://superdarn.thayer.dartmouth.edu/aacgm.html> [Accessed 26 Apr. 2025].
- Vorobjev, Gustafsson, G., G.V. Starkov, Feldstein, Y.I. and N.F. Shevnina (1975). Dynamics of day and night aurora during substorms. *Planetary and Space Science*, [online] 23(2), pp.269–278. doi:[https://doi.org/10.1016/0032-0633\(75\)90132-4](https://doi.org/10.1016/0032-0633(75)90132-4).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017). Attention Is All You Need. [online] arXiv. Available at: <https://arxiv.org/abs/1706.03762>.
- Baumjohann and Treumann, R.A. (2012). Basic Space Plasma Physics. IMPERIAL COLLEGE PRESS eBooks. doi:<https://doi.org/10.1142/p850>.
- Zhang, Q., Wang, Y. and Wang, Y. (2023). On the Generalization of Multi-modal Contrastive Learning. [online] arXiv.org. Available at: <https://arxiv.org/abs/2306.04272>.
- Zhang, X., Alavi, A., Li, M. and Liu, Z. (2022). Self-Supervised Contrastive Learning for Medical Time Series: A Systematic Review. *Sensors*, [online] 23(9), p.4221. doi:<https://doi.org/10.3390/s23094221>.
- Akasofu (1964). The development of the auroral substorm. 12(4), pp.273–282. doi:[https://doi.org/10.1016/0032-0633\(64\)90151-5](https://doi.org/10.1016/0032-0633(64)90151-5).
- Eastwood, J.P., Nakamura, R., L. Turc, L. Mejnertsen and Hesse, M. (2017). The Scientific Foundations of Forecasting Magnetospheric Space Weather. [online] doi:https://doi.org/10.1007/978-94-024-1588-9_10.

The code used for this study can be found on the following GitHub repository.