Imperial College London

Department of Earth Science and Engineering

MSc in Applied Computational Science and Engineering

Independent Research Project
Final Report

Using deep learning to predict water ice presence at the lunar poles

by

Ardan Suphi

Email: ardan.suphi23@imperial.ac.uk
GitHub username: acse-as5023
Repository: https://github.com/ese-msc-2023/irp-as5023

Supervisors:

Dr Kathryn Hadler
Dr Philippa Mason

August 2024

Contents

Abstract

Recent measurements, both in-situ and from remote sensing, have indicated large quantities of cold-trapped volatile materials, such as water ice, on the lunar surface. Many believe water ice to be contained in permanently shadowed regions (PSRs) and remote sensing has been used to estimate where water ice is most likely to be found. This report uses machine learning to interpret several remote sensing datasets and score any point at the lunar poles (latitudes below -75° and above 75°), not just in PSRs, for its resource potential. NASA's Diviner, LOLA, M^3 and Mini-RF datasets were collected, combined and synthetically labelled to create features for machine learning models. A fully connected neural network (FCNN) and a graph convolutional network (GCN), both with self-attention and a residual connection, were developed and trained to predict a score between 0 and 7, with 7 representing the highest indication of water ice. The FCNN performed much better than the GCN, producing predictions with coherence and relevance to the test data. The GCN performed relatively poorly and struggled to capture trends within the data.

1 Introduction

1.1 Problem

The presence, location, and abundance of volatile substances, such as water ice, on the lunar surface have been the subject of long-standing scientific inquiry with many considering the most likely locations for them as inside permanently shadowed regions (PSRs). Individual remote sensing datasets have yielded interesting insights and when considered together, can offer indications of where water ice is most likely.

This report proposes that machine learning (ML) can be used to identify areas of high water ice potential using fewer datasets. It suggests that a data driven approach, considering entire datasets, can predict the likelihood of water ice presence for any point at the lunar poles, not just points within PSRs. By more efficiently using each dataset, future remote sensing missions can be more targeted in terms of resolution and location.

1.2 Objectives and Expected Deliverables

The objectives of this project were to collect, interpolate, combine and label several remote sensing datasets, then train a machine learning model to score coordinate points on the lunar surface for the presence of volatiles, based on remote sensing indicators.

As inputs, the model will receive coordinates plus several remote sensing values and output a score between 0 and 7 with the former representing no indication of volatile material and the later representing the maximum possible indication of such materials.

1.3 Literature Review Summary

Literature which evaluates or ranks points on the lunar surface on a large scale seems scarce. Several have achieved this analytically ([1, 2, 3]) but current efforts in machine learning (as applied to lunar science) appear more directed at specific capabilities such as image enhancement, rover navigation and building models around specific datasets such as Narrow Angle Camera (on board the Lunar Reconnaissance Orbiter) images [4] and LOLA [5].

Brown et al. [1] combined 8 remote sensing datasets and assigned scores to rank water ice likelihood in PSRs. This was done by trimming datasets from the Planetary Data System to include only permanently shadowed areas and using geological rationale to assign labels. This project attempts to use only 4 of these datasets to evaluate the entire region

Another brief example of similar work is when Chen et al. [6] applied unsupervised learning to five morphometric (diameter, area, etc.) and two elemental (FeO and TiO_2 contents) features to find which combination is representative of the properties of lunar domes. This work successfully gathered several remotely available features and trains on numerical values.

More broadly, Wu et al. [7] discuss the opportunities which are arising as remote sensing and computational intelligence (CI) become more integrated.

Following a wider review, by far the most common deep learning architecture was the convolutional neural network (CNN), commonly used to process images, however other architectures feature in special cases such as siamese neural networks [8], reinforcement learning [9], value iteration [10] and more. There are also opportunities for current research areas to be combined. For example, models which denoise PSR images and those which allow rovers to navigate autonomously in lit regions could be combined to enable rovers to traverse into PSRs.

1.4 Datasets

Three of the datasets used in this project are generated by NASA instruments onboard NASA's Lunar Reconaisance Orbiter (LRO): Diviner, LOLA and Mini-RF, and one is from a NASA instrument which was onboard ISRO's Chandrayaan-1: M^3 before Chandrayaan-1's failure in 2009 [11]. More detailed descriptions of the datasets including how they were accessed, and a list of datasets included in Brown et al. [1] but omitted in this project, can be found in Appendix A2.

All datasets are publicly available from the Geosciences Node of NASA's Planetary Data System (PDS) [12], including all required metadata, documentation and instructions for use. Table 1 illustrates how each remote sensing dataset is used to indicate the presence of water ice. Data preprocessing and product selection were also important factors and explained in Appendix A3 with data storage outlined in Appendix A4.

Table 1: Criteria used for scoring resource potential from each dataset in the PSR volatiles ranking. Modified from Brown et al. [1]

Dataset	Strongly Consistent (2)	Consistent (1)	Inconsistent / No Detection (0)	Notes
Diviner	<110 K	n/a	>110 K	Average temperature. Sublimation of water ice >110 K
LOLA	Observations consistent with water ice cover >3 km^2 in PSR	Observations consistent with water ice cover <3 km^2 in PSR	No observations consistent with water ice	n/a
M^3	Ice detections cover >2.4 km^2 in PSR	Ice detections cover <2.4 km^2 in PSR	No near-IR ice detection	n/a
Mini-RF	n/a	CPR-anomalous	Rough, multi-bounce, or no enhanced radar signature	Interpretation of water ice within the upper 2 to 3 m of regolith

2 Computational Overview

This project developed standalone code from scratch. All code was written in Python with PBS shell scripts provided as this code was regularly computed on Imperial's High-Performance Computing (HPC) clusters. An `environment.yml` file is also provided for those using Conda and all dependencies can be found in the `requirements.txt` file. Further details are available in the `README` section of this project's GitHub repository.

3 Early investigations

This work seeks to develop a labelling model using fewer datasets than Brown et al. [1] and as such, an investigation was done into the impact of removing some such datasets. Kendall's Tau and Spearman's Rank Correlation are both measures of strength and direction of association between ordinal rankings. Justification for selecting these metrics can be found in Appendix A6. The results, comparing a ranked list of the 4 selected datasets to all 8 were:

- Kendall's Tau: 0.8816 (p-value: 1.46e-18)

- Spearman's Rank Correlation: 0.9426 (p-value: 1.06e-31)

Both of these metrics suggest strong correlation between the two lists and that the impact of removing the selected datasets would not significantly affect the outcome of this project.

Additionally, Rank-Biased Overlap (RBO) was applied and this result was much lower (0.3902). RBO measures the similarity between the top ranks of the two lists, putting more weight on the top items. If the top rankings are not very similar but the overall order of the lists is similar, this would explain why Kendall's Tau and Spearman's Rank Correlation show high values and RBO does not.

4 Data Processing

All 4 datasets were run through the general workflows described in Figures 1a and 1b however each dataset presented unique challenges which were addressed at each stage. This section discusses the process to obtain, decode and process the data, then interpolate, combine and synthetically label it. This process is designed to create a list of feature-label pairs required as inputs to the models.

4.1 Reading in encoded data

For each dataset, either a URL or local path was defined to a directory containing the image and label files. The number of files for each dataset is seen in Table 4 in Appendix A5. A `Dask` client is then set up to allow files to be processed in parallel and the `asyncio` module is used for asynchronous processing. A wrapper script creates 12 CSV files for longitude range in 30 degree increments (0-30, 30-60, ..., 330-360). The data was saved in this way for better memory management and to allow for parallel processing of the data. Metadata, image file urls and supporting information were submitted to the core image processing function.
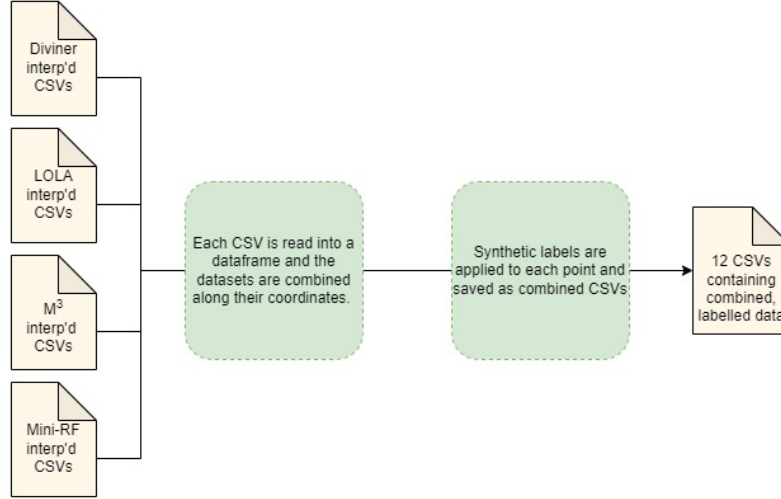
4.2 Image processing methodology

Image processing methodology is aligned across all datasets but separate functions for the LRO datasets and M^3 were developed due to different file structures. The image processing steps are:

1. Extract the image

2. Process the image

3. Generate the coordinates
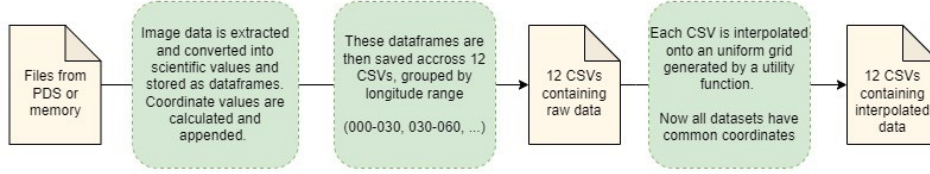
(a) Data processing pipeline for each individual dataset.

(b) Data processing pipeline to combine datasets.

Figure 1: Data processing pipelines. (a) is applied four times, to each dataset and following that, (b) is used to bring them together.

Following these steps, the dataframe is constructed and populated, several checks and optimisation steps are completed and the final dataframe is returned to the wrapper script.

4.3 Processing LRO images

For URLs, the `requests` library was used to obtain image data from the web. Since the Mini-RF dataset is packaged into a single file, the metadata variables `lines` and `line_samples` (used determine the 2D matrix shape) can be used to split the binary data cleanly, enabling partial extraction of scientific values without decoding the entire file.

For images with the .jp2 extension (Diviner, LOLA), the `glymur` library is used to decode the data. For images with the .img or extension (Mini-RF), metadata variables are used to convert the binary data into scientific values manually.

The metadata specifies a 'missing value', which was replaced with 'Not-a-Number' (NaN) indicators and an affine transformation was applied using the scaling factor and offset from the metadata. A final check ensures all values are within range before returning output values. Coordinates for LRO datasets are generated according to a projection type, equations for which can be found in Appendix A7.

4.4   Processing M³ images

The first new challenge with the M³ data is that it is 3D, providing values at 86 spectral bands for each point. Calibration files [13] provided the wavelengths of each band and Brown et al. [1] specifies the target wavelengths for water ice reflectance are 1.3, 1.5 and 2.0μm. The values at these bands are kept, as well as the values at one band either side of the target, resulting in 9 out of the 86 bands returned as image data. Missing values were again replaced with NaNs but this time no transformation was required.

The raw values from this dataset, however, are not the desired indicators and, unlike for LRO datasets, must be processed further. The now decoded data contains reflectance values of light at given wavelengths and will be converted into band depth ratio (BDR) values to give a better indication of the presence of volatile materials. The two reference bands (one higher and lower) for each target band are known as shoulders and the target band itself is known as a trough. The BDR value at a given point and wavelength is defined in Equation 1.

$$\text{BDR} = \frac{S_{\text{h}} + S_{\text{l}}}{2 \times T} \tag{1}$$

Band Depth Ratio Equation

S_u : Reflectance at higher shoulder $\qquad\qquad$ T : Reflectance at trough
S_l : Reflectance at lower shoulder

If water ice is present, T is expected to increase sharply while the shoulder values would remain constant, resulting in a dip in BDR values and indication of water ice. As 3 target wavelengths are present, each point has 3 BDR values and so the minimum value was taken at each point and a clip of 1.75 was applied to outliers. These BDRs are now the output values.

Finally, for the M³ coordinates, image and label files for the location data were retrieved and decoded similarly to the main image file, no projections were used. The location image, however, contained 3 bands. For each point, a latitude and longitude value was present but included in this data was LOLA elevation data, defined as the distance from the given point to the centre of the Moon. This was accepted as an additional parameter for use in the models later. The radius values were subtracted from a NASA reference radius [14] to obtain elevation values and returned to the main function alongside coordinates.

4.5   Interpolation

Once each of the datasets had been converted to CSVs, the next stage of the data processing was to interpolate each dataset onto a common coordinate mesh. Brown et al. [1] considered PSRs in the latitude ranges of $[-90, -75]$ and $[75, 90]$, which is repeated here. A resolution of 240m was used as this closely aligned with the resolutions of the datasets, seen in Table 4. A finer initial mesh was preferred as the resolution can be artificially decreased later on, but not increased without additional interpolation.

The interpolation method selected was the inverse distance weighting method which assumes points closer to the location of interest have a stronger influence on the interpolated value than those farther away. The interpolation algorithm begins by using a KDTree (from `scipy.spatial`) to identify neighbours of the new point within a certain radius and interpolate based on these points. The `concurrent.futures.ProcessPoolExecutor` package was used to interpolate files in parallel and save them into the same longitude ranges.
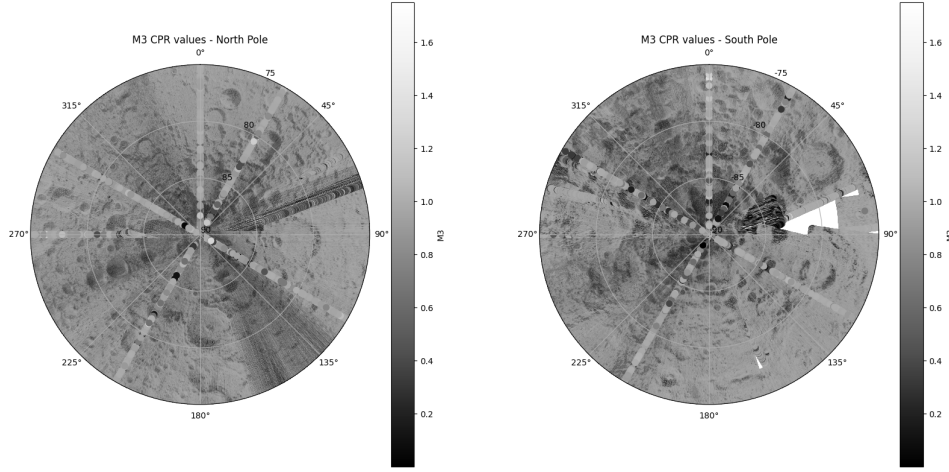
(a) Map of raw M³ values
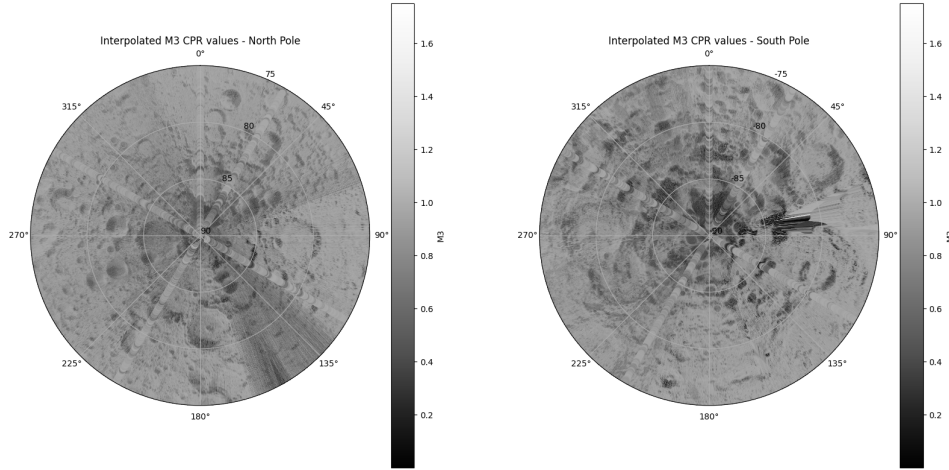
(b) Map of interpolated M³ values

Figure 2: The raw M³ data versus the interpolated data. Important features can be seen across both and patches where Chandrayaan-1 wasn't able to fly over are filled in.

4.6 Labelling

Interpolated data from each dataset is combined into a single dataframe along the common coordinates. Each dataset is labelled 0-2 and these labels are summed so that each point receives an overall score. The criteria used in this project can be seen in Table 1, modified from Brown et al. [1].

The 'consistent' value for Diviner was excluded due to its basis on seasonal criteria and no 'strongly consistent' rank is provided for Mini-RF as the monostatic radar cannot discriminate between blocky materials and water ice in the subsurface [15].

A clear criteria for 'ice detections' for M³ and 'CPR-anomalous' for Mini-RF were not discussed in Brown et al. [1]. Observations consistent with water ice for LOLA was described as those 2 times the

standard deviation (σ) above the mean (μ) and so this method was applied to all relevant datasets. For LOLA and Mini-RF, the values above $\mu + 2\sigma$ were taken as indicative, and for M^3, the values below $\mu - 2\sigma$ were taken as indicative.

The LOLA and M^3 datasets also have a spatial aspect to be considered, seen in Table 1. To determine the area, a mask was used to convert the data into binary array where 1 represented a value above the threshold. The `scipy.ndimage.binary_dilation` package was used to dilate the array using a 1-wide disk and `label` from the same library was used to identify features within it. Dilation was used as a very large portion of the points were not connected to any others and the features which did contain multiple points contained very few. This was considered unrealistic as many PSRs are larger than the 0.06 km^2 resolution of the interpolated grid. The area of each feature was calculated and points assigned a value of 1 or 2 depending on the size of the feature they belonged to. Labels were cumulatively applied for each dataset and the final dataframes were saved as combined and labelled CSVs. An illustration of the labelled points can be seen in Figure 3.

Side note: The binary arrays produced some interesting images, included in Appendix A10 for interest.
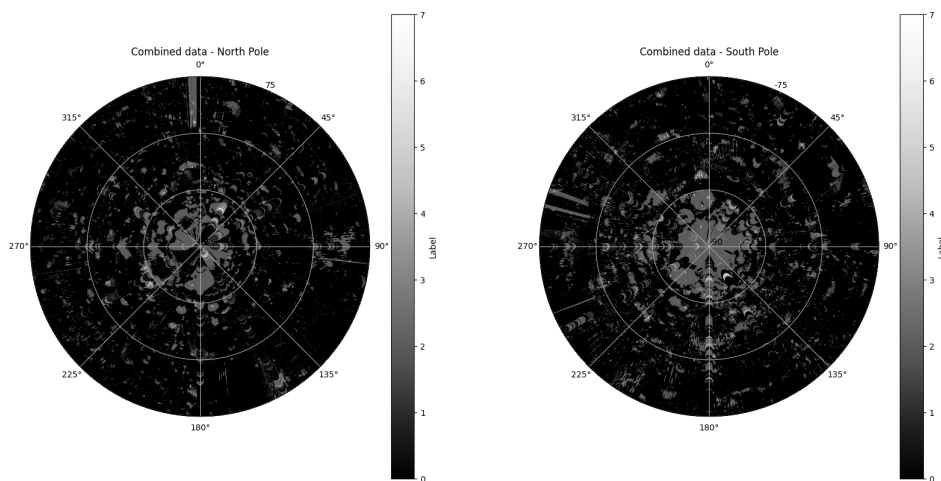

Figure 3: A heatmap of labels for the combined datasets at each pole

4.7 Validation and Testing

As many of these programs are designed to be run once, continuous integration (CI) tools such as GitHub Actions were not used. Instead, exception handling mechanisms are implemented throughout the functions to ensure that at each stage, outputs are as expected and only the correct data is flowing through the pipeline.

For image data, examples of checks that are in place include for file type, data type, image size and projection type. All of these values, are confirmed to guarantee a correct dataframe at the end. If any of these checks are not passed, exception handling is used to identify the issue. This was assisted by the metadata in some cases, such as Diviner having a 'derived maximum'. Some of these checks simply removed incorrect values, rather than raising an error. For data processing, all processes can also be verified through visual inspection of the generated images such as Figures 2 and 3.

5 Machine Learning

Two ML models were developed to train on the labelled data generated above: a fully connected neural network (FCNN) and a graph convolutional network (GCN). Both had self-attention mechanisms and residual layers built in. They were developed independently with regularisation and hyperparameter tuning being separate to try and develop the best model from each architecture. The input data consists of seven features: two spatial coordinates (latitude and longitude) and five remote sensing values (Diviner, LOLA albedo, LOLA elevation, M^3 and Mini-RF). The target label represents a continuous score indicating how consistently the datasets indicate the presence of water.

The training data was augmented as the original dataset was extremely biased. Oversampling was used to increase the representation of some labels and undersampling to decrease the representation of others. Given the majority of points are not expected to indicate water ice presence, a bias towards a label of 0 was retained. Table 2 contains the original and target proportions of the data. In some cases, where this proportion could not be met for a specific label, the values were made up with the second least represented value.

Table 2: Proportions of original and augmented data

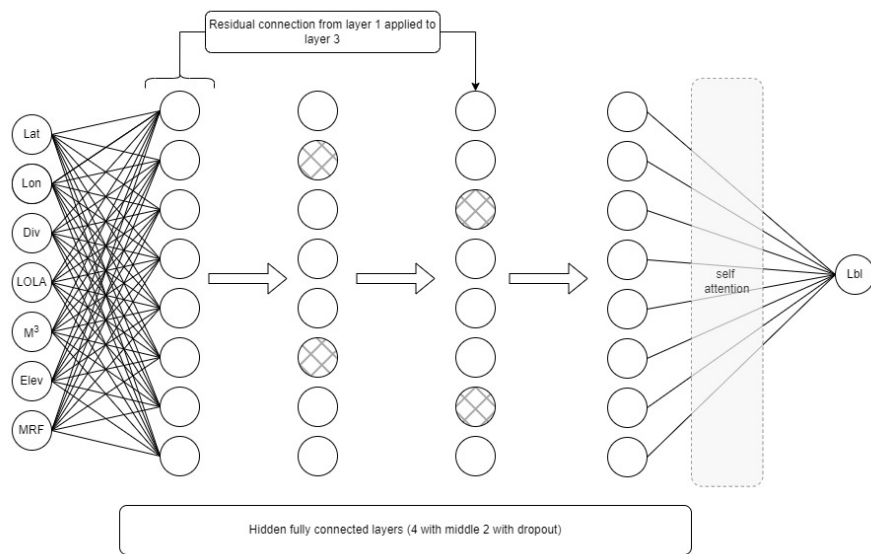| Label	Original labeled data (%)	Target for training data (%)
0	60.92	22.22
1	6.74	11.11
2	27.00	11.11
3	3.16	11.11
4	1.98	11.11
5	0.15	11.11
6	0.04	11.11
7	<0.01	11.11

5.1 Fully Connected Neural Network (FCNN)

Figure 4: FCNN model architecture

The FCNN architecture begins with an initial sequence of fully connected (FC) layers, each followed

by batch normalization and ReLU (Rectified Linear Unit) activation function, which are critical for stabilising training and accelerating convergence. The core of the network includes a self-attention mechanism that allows the model to focus on the most relevant features or combinations of features, thus improving its ability to learn patterns indicative of water ice presence. This self-attention layer dynamically adjusts the importance of different features by computing attention scores, which are then used to reweight the features before passing them to the subsequent layers. For the FCNN, a custom self-attention class was developed and implemented to be included as a layer in the model.

To further enhance the model's learning capacity, residual connections are employed. Specifically, after the first fully connected layer, the output is stored and later added back to the transformed feature map from subsequent layers. This residual connection improves gradient flow during backpropagation, allowing the model to train effectively and capture more complex patterns within the data. The model is regularized using dropout after key layers to prevent overfitting, and the final output layer generates the predicted water ice score.

The model was trained using the Adam optimizer and all features were standardised and normalised before being loaded into the model. The loss function used was Smooth L1 loss, also known as Huber loss.

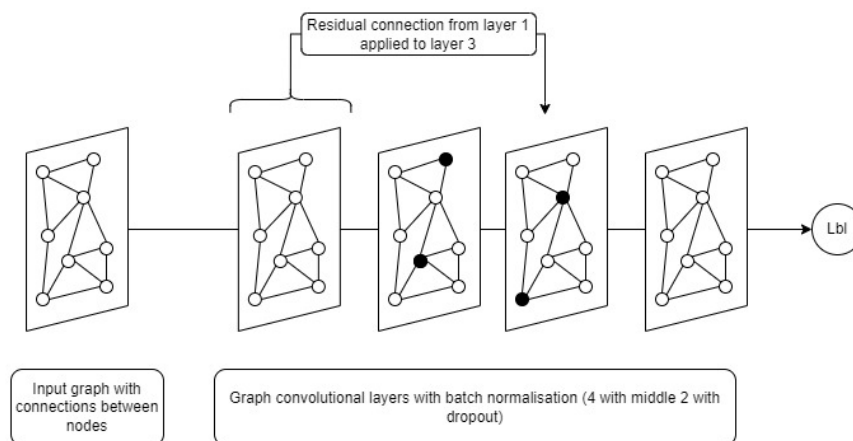## 5.2   Graph Convolutional Network (GCN)

Figure 5: GCN model architecture

A Graph Convolutional Network (GCN) was developed to leverage the spatial structure of the lunar surface data. In this model, each data point (node) is represented in a graph structure where edges connect nodes based on spatial proximity or similarity in remote sensing values. This graph representation allows the GCN to capture spatial relationships and dependencies that may be critical for accurately predicting water ice presence.

The GCN architecture consists of multiple graph convolutional layers, each followed by batch normalization and ReLU activation functions. These layers aggregate and refine features from neighboring nodes, effectively capturing both local and global spatial dependencies. Similar to the FCNN, residual connections are utilized within the GCN to preserve the original feature information and improve training stability, and self-attention to create preference for different features using a multi-head attention mechanism from `nn.MultiheadAttention`. After the initial graph convolution, the output is stored and added back to the feature map after the subsequent graph convolutional layers, enhancing the model's ability to capture complex spatial patterns.

K-Nearest Neighbors (KNN) was chosen over a fully connected GCN to efficiently capture local

dependencies with lower computational complexity. A fully connected GCN can capture global dependencies by connecting each node to all others, but comes with high computational and memory costs.

The GCN model is regularized using dropout after each graph convolutional layer, and the Adam optimizers used for training. Features were standardised and normalised for the GCN and the loss function used was also Huber loss.

5.3 Running the models

After training, the best models were saved as .pth files and the scalars for normalising and standardising input data were saved using `joblib`. The models and scalars were loaded into a separate file and the models inferred a label from a random input from the combined data. The FCNN output a single label per input and the GCN output a graph of labels for each input graph. Single nodes could be extracted from the GCN for comparison.

5.4 Considerations and Potential Improvements

Given the large datasets, several improvements were made to the code to improve computational efficiency. Firstly, the dataloaders were used to pin memory if GPUs were being used and the prefetch factor preloaded batches into memory. Torch's grad scaler was used to implement mixed precision training and the models were trained on a reduced dataset size.

Potential alternatives or enhancements include hybrid models that combine the strengths of both FCNNs and GCNs, such as a model that uses a GCN to preprocess spatial data before passing it to an FCNN with self-attention. With additional time, a transformer architecture may also have been considered, as well as the CNN architecture.

5.5 Validation and Testing

The data was split into training (80%), validation (10%) and test (10%) sets and a stratified shuffle split was used to ensure that each set contained a proportional number of points with each label. The validation set was used to compare model training to unseen data in real time and the test set was only used to evaluate the final model. Where standardisation and normalisation was used, this was fit to the train set only.

Both models considered various loss functions, including Huber loss, mean squared error (MSE), and mean absolute error (MAE). Due to time constraints, a comprehensive comparison wasn't possible, so Huber loss was chosen for its robustness, balancing MSE's sensitivity to precision and MAE's stability against outliers. During hyperparameter tuning, the Huber loss 'beta' parameter, which determines the transition point between MSE and MAE, was optimized.

6 Results

Using the augmented data, the training results for the FCNN can be seen in Figure 6 with a final test loss of 0.5328. The results for the GCN can be seen in Figure 7 with a final test loss of 1.2303. The saved models were applied to 10 randomly selected points to characterise the losses. The results can be seen in Table 3 and the hyperparameters used in the final version of both models can be found in Appendix A8.
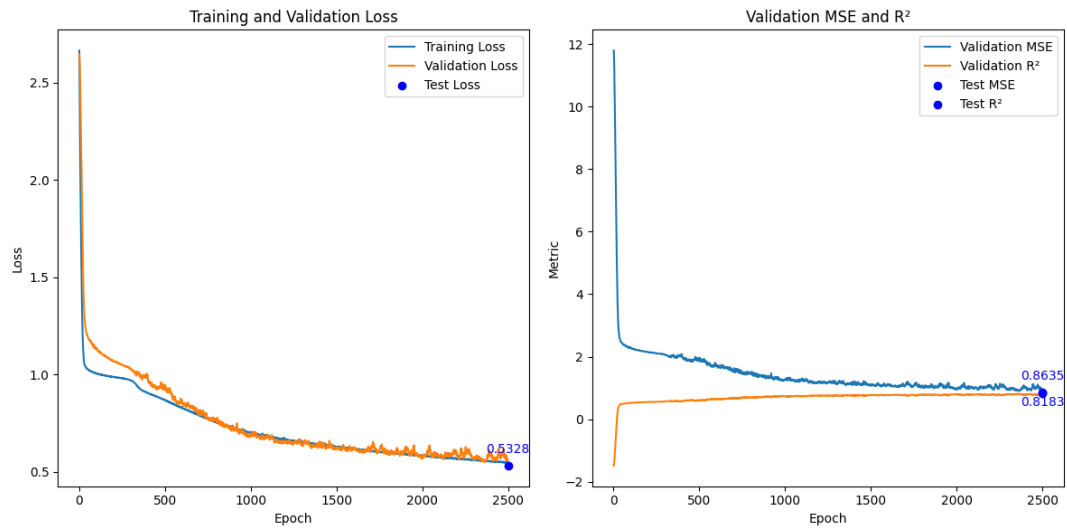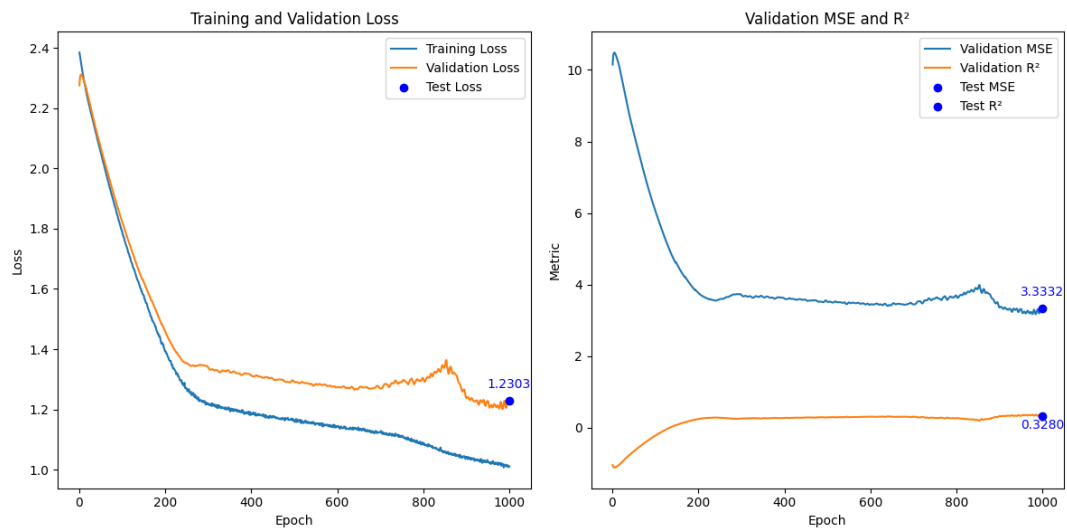
Figure 6: Training results for FCNN

Figure 7: Training results for GCN

Table 3: Final models applied to randomly selected points

(a) Predictions of FCNN

Trial	Latitude	Longitude	True label	Predicted label
1	77.64	88.57	0	0.065
2	88.07	63.61	1	1.093
3	-87.05	131.78	2	2.361
4	87.13	132.07	3	3.331
5	82.72	202.15	4	3.972
6	-82.88	266.33	5	5.017
7	-81.74	147.86	6	6.174
8	-86.07	246.92	7	5.331
9	-84.84	37.27	1	-0.026
10	-82.91	302.84	0	0.007

(b) Predictions of GCN

Trial	Latitude	Longitude	True label	Predicted label
1	-76.51	270.51	0	-1.29
2	76.27	50.98	1	-0.31
3	88.68	334.95	2	2.46
4	-83.87	0.25	3	0.70
5	83.67	320.28	4	1.62
6	-88.94	322.15	5	2.74
7	84.80	252.03	6	5.62
8	-89.85	150.04	7	4.55
9	79.84	244.55	1	0.13
10	-83.86	37.95	4	1.48

7 Discussion

7.1 Early investigations

Kendall's Tau, Spearman's Rank Correlation and RBO were used to quantify the effect on removing four of the eight datasets from Brown et al. [1]. Given the high values for the first two metrics and lower value for the third, it can be inferred that for PSRs with low resource potential, fewer datasets are required to relegate them to a lower position on a rankings table, however, for PSRs with a higher potential to contain volatile materials, a larger number of datasets are needed to distinguish between those top few. An investigation into whether machine learning can better distinguish PSRs at the top of the list without the need for additional datasets then becomes more important.

7.2 Data processing

When condensing the Band Depth Ratio (BDR) values during M^3 image processing, the minimum value of each was taken. Due to the division by a low decimal number, there was significant variance in the maximum BDR values making it difficult to plot and to clip, a similar issue was seen with the mean values. Plotting the minimum, however, allowed regions with spikes to be highlighted without being clouded by extremely high values. A clip of 1.75 was selected by investigating the percentiles and choosing a value far beyond the 99th percentile. It was validated by subsequent analysis of the number of points that were clipped.

The method of interpolation, inverse distance weighting (IDW), is a common but simple method for interpolating datapoints onto a grid. More advanced techniques such as kriging and radial basis function interpolation were considered but deemed too complex for the time available during this project. The optimum interpolation function would capture clear boundaries within the data (such as PSR borders at crater walls).

For the synthetic labelling, Figure 3 shows a clear increase in scores closer to the poles and especially elevated scores at the lunar south pole aligning with current theory that the south pole is most likely to contain volatile materials [16]. Patches can be seen in these regions but not the iconic circular crater sites seen in Figure 2. This is likely because the shadowed regions have been captured, which take different shapes to the craters themselves. A consideration, however, is the impact of the missing data in the Diviner dataset. Appendix A9 shows the raw and interpolated images for Diviner and it can be seen that the the interpolation of the missing centre infers a low average temperature across all of those points. These interpolated points are likely to have been scored highly and contributed to higher labels around the poles. The impact of this is, however, limited as the labels in these regions can clearly be seen to be mid to high range and so must be scoring highly across several more, if not all, datasets. As the only dataset which did not have a criterion for a label of 1, and the likelihood of a larger number of points having a low average temperature compared to those likely to contain water, Diviner is likely to contribute to the high proportion of points labelled 2 seen in Table 2.

During the dilation of the LOLA and M^3 datasets for the area calculation, it was observed that the number of features for M^3 was significantly larger than for LOLA. This means that the LOLA points originally above threshold were much more clustered and those of M^3 were much more dispersed. This resulted in the dilation process creating more new points labelled 1 or 2 for M^3 compared to LOLA. The dilation process was designed to connect points which were nearby but not adjacent and this appears to have succeeded for LOLA but the M^3 values above threshold were unable to aggregate with this level of dilation. The shape wasn't increased as this would have caused too many points to be designated above threshold when the numerical value was not.

7.3 Machine learning and results

The FCNN architecture was selected as it was a good baseline model. FCNNs are straightforward to train, effective at extracting high-dimensional features and are flexible as they do not assume any specific structure of the input features. The primary drawback of this selection was the lack of spatial awareness. FCNNs treat all input features independently and so are unable to capture the geospatial patterns which may be inherent in the data.

The GCN architecture was selected as a more ambitious model as it is designed to handle data structured as graphs, making them excellent for modelling spatial relationships directly. Nodes in GCNs aggregate information from their neighbours which can be particularly useful in geospatial tasks. This model architecture is certainly more complex and nuanced and so data preprocessing, memory requirements and computation time are all important factors

The results show that the FCNN performs better than the GCN, showing lower test and validation losses. This may be because the labelling process was done using four features, and not coordinate data, and so if coordinate data was included in the labelling process (i.e. scoring points closer to the poles or closer to the Moon's far side higher), the GCN may see improvement. This contrast is likely exacerbated by the self-attention mechanism in the FCNN. The increased complexity of the GCN is also likely to be a factor and additional data preparation, model tuning and theoretical application may have enabled the GCN to perform better.

The GCN can be seen to significantly overfit and was not trained on as many epochs as the FCNN for this reason. An aggressive approach was taken to try and address this overfitting with a final dropout rate of 0.75 and weight decay of 0.05 compared to 0.3 and 7.5e-4 for the FCNN. This reinforces the argument that model architecture was unsuitable and improved theoretical application and model development was necessary to enable the model to generalise.

When training on the raw data (i.e. before augmentation by resampling), both models achieved lower losses compared to those in Section 6 however when examining those predictions, both models predicted either 0 or 7 exclusively, with the GCN predicting a label of 0 for every input. Given that less than 3% of the lunar surface is permanently shadowed [17] and less than 6% of the original data was labelled 3 or higher, the models had likely learned that if they predict a 0, they are usually right. To improve these estimates, the training data was resampled, as seen in Table 2, and while the test losses increased for both models, the predictions can be seen to align much more closely with the true labels.

For the FCNN, many of the predicted labels can be seen to align with the true labels with some variance observed due to the regression task, as opposed to classification. Some interesting interpretations can be seen in these results where, for example, the label for trial 4 was increased slightly from a true value of 3 to a predicted 3.331 whereas the true value at trial 5 was decreased in the prediction from 4 to 3.972. The model has made larger decisions too, such as in trial 9 where a true label of 1 was predicted as -0.026, which can be approximated to 0. These results, in addition to 6 out of 10 trials being predicted within a range 0.2 of their true values indicate that the model may be understanding trends within the data and actively changing the labels of selected points. While the trials with true labels of 5 and 6 were predicted fairly accurately, trial 8, trying to predict a true label of 7, struggled and only predicted 5.331. This is likely due to the lack of points labelled 7 and may indicate the network struggling to model the highest label.

The predictions for the GCN were much less aligned, as expected from the higher loss. There appeared to be a general trend with the high labels being predicted highly and lower labels predicted as lower however large deviances could be seen. No points were within 0.2 of the true value and most concerning result was trial 1 which predicted a label of -1.29 (true label of 0). This indicates the model struggling to learn the basic range of labels and so is highly unlikely to have captured deeper trends.

These models use an ML methodology to provide a single label to any point on the lunar surface. Those directing the search for water on the moon now have a computational perspective to go alongside the geoscientist's perspective put forward by Brown et al. [1] and others. More advanced model architectures, lower level data products and more complex interpolation techniques could be used to enhance further the predictions of an ML algorithm. New datasets from missions such as NASA's Lunar Trailblazer can also be included to enhance the results. Ultimately, this research intends to make a small contribution to the seismic task of finding and understanding the nature of water ice on the Moon.

8 Conclusion

Following the successful processing and interpolation of Diviner, LOLA albedo, M^3, Mini-RF and LOLA elevation and labelling according to the first 4 of these datasets, features were created to train a fully connected neural network (FCNN) and graph convolutional network (GCN). The data was resampled, preprocessed and used to train these models. The FCNN performed well with a test loss of 0.5328 and a clear coherence amongst predictions of randomly selected test points. The GCN, however, performed poorly with a much higher error, 1.2303, a significant amount of overfitting and a model which struggled to capture trends. This project successfully developed a well-performing model able to predict the likelihood of water ice presence for any point on the lunar surface.

References

[1] H.M. Brown et al. "Resource potential of lunar permanently shadowed regions". In: *Icarus* 377 (2022), p. 114874. ISSN: 0019-1035. DOI: https://doi.org/10.1016/j.icarus.2021.114874. URL: https://www.sciencedirect.com/science/article/pii/S001910352100511X.

[2] H.M. Brown, M.S. Robinson, and A.K. Boyd. "Identifying Resource-rich Lunar Permanently Shadowed Regions". In: *Developing a New Space Economy*. Temple, AZ, 2019. URL: https://www.hou.usra.edu/meetings/lunarisru2019/pdf/5035.pdf.

[3] Yutong Jia et al. "Selection of Lunar South Pole Landing Site Based on Constructing and Analyzing Fuzzy Cognitive Maps". In: *Remote Sensing* 14.19 (2022), p. 4863.

[4] Ben Moseley et al. "Extreme low-light environment-driven image denoising over permanently shadowed lunar regions with a physical noise model". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 6317–6327.

[5] Michael K Barker et al. "Improved LOLA elevation maps for south pole landing sites: Error estimates and their impact on illumination conditions". In: *Planetary and Space Science* 203 (2021), p. 105119.

[6] Yuchao Chen et al. "Unsupervised machine learning on domes in the lunar gardner region: implications for dome classification and local magmatic activities on the moon". In: *Remote Sensing* 13.5 (2021), p. 845.

[7] Yue Wu et al. *Computational Intelligence in Remote Sensing*. 2023.

[8] Benjamin Wu et al. "Absolute localization through orbital maps and surface perspective imagery: A synthetic lunar dataset and neural network approach". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 3262–3267.

[9] Max Pflueger, Ali Agha, and Gaurav S Sukhatme. "Rover-IRL: Inverse reinforcement learning with soft value iteration networks for planetary rover path planning". In: *IEEE Robotics and Automation Letters* 4.2 (2019), pp. 1387–1394.

[10] Brian Gaudet, Richard Linares, and Roberto Furfaro. "Deep reinforcement learning for six degree-of-freedom planetary landing". In: *Advances in Space Research* 65.7 (2020), pp. 1723–1741.

[11] Indian Space Research Organisation. *Chandrayaan-1*. https://www.isro.gov.in/Chandrayaan_1.html. Accessed: 2023-08-16. 2023.

[12] NASA Planetary Data System Geosciences Node. *Planetary Data System Geosciences Node*. https://pds-geosciences.wustl.edu/. Accessed: 2024-08-12. 2024. URL: https://pds-geosciences.wustl.edu/.

[13] NASA Planetary Data System. *Chandrayaan-1 M3 calibration files*. https://planetarydata.jpl.nasa.gov/img/data/m3/CH1M3_0003/CALIB/. Accessed: 2023-08-23. 2024.

[14] Peter Isaacson et al. *M3 Data Tutorial*. Workshop presentation. Accessed: 2024-08-15. Nov. 2011. URL: https://pds-imaging.jpl.nasa.gov/documentation/Isaacson_M3_Workshop_Final.pdf.

[15] PD Spudis et al. "Evidence for water ice on the Moon: Results for anomalous polar craters from the LRO Mini-RF imaging radar". In: *Journal of Geophysical Research: Planets* 118.10 (2013), pp. 2016–2029.

[16] Paul O Hayne et al. "Evidence for exposed water ice in the Moon's south polar regions from Lunar Reconnaissance Orbiter ultraviolet albedo and temperature measurements". In: *Icarus* 255 (2015), pp. 58–69.

[17] J Andrew McGovern et al. "Mapping and characterization of non-polar permanent shadows on the lunar surface". In: *Icarus* 223.1 (2013), pp. 566–581.

[18] NASA Planetary Data System. *LRO Diviner Gridded (Map) Data*. https://pds-geosciences.wustl.edu/lro/urn-nasa-pds-lro_diviner_derived1/. Accessed: 2023-08-16. 2024. DOI: 10.17189/1520647.

[19] NASA Jet Propulsion Laboratory. *Diviner Lunar Radiometer Experiment (DLRE)*. `https://www.jpl.nasa.gov/missions/diviner-lunar-radiometer-experiment-dlre`. Accessed: 2024-08-12. 2024. URL: `https://www.jpl.nasa.gov/missions/diviner-lunar-radiometer-experiment-dlre`.

[20] NASA Planetary Data System. *LRO Lunar Orbiter Laser Altimeter Gridded Digital Elevation Models*. `https://pds-geosciences.wustl.edu/lro/lro-l-lola-3-rdr-v1/lrolol_1xxx/data/lola_gdr/`. Accessed: 2023-08-16. 2024. DOI: `10.17189/1520642`.

[21] David E. Smith and Maria T. Zuber. *Lunar Orbiter Laser Altimeter (LOLA) Fact Sheet*. Tech. rep. Accessed: 2024-08-12. Goddard Space Flight Center, NASA, 2007. URL: `https://your-source-link-here`.

[22] Elizabeth A Fisher et al. "Evidence for surface water ice in the lunar polar regions using reflectance measurements from the Lunar Orbiter Laser Altimeter and temperature measurements from the Diviner Lunar Radiometer Experiment". In: *Icarus* 292 (2017), pp. 74–85.

[23] NASA Planetary Data System. *Chandrayaan-1 M3 dataset*. `https://planetarydata.jpl.nasa.gov/img/data/m3/`. Accessed: 2023-08-16. 2024.

[24] NASA Jet Propulsion Laboratory. *Moon Mineralogy Mapper (M3)*. `https://www.jpl.nasa.gov/missions/moon-mineralogy-mapper-m3`. Accessed: 2024-08-12. 2024.

[25] RO Green et al. "The Moon Mineralogy Mapper (M3) imaging spectrometer for lunar science: Instrument description, calibration, on-orbit measurements, science data calibration and on-orbit validation". In: *Journal of Geophysical Research: Planets* 116.E10 (2011).

[26] NASA Planetary Data System. *LRO Mini-RF Global Mosaic Data*. `https://pds-geosciences.wustl.edu/lro/lro-l-mrflro-5-global-mosaic-v1/`. Accessed: 2023-08-16. 2024. DOI: `10.17189/1520656`.

[27] Joshua TS Cahill et al. "The miniature radio frequency instrument's (Mini-RF) global observations of Earth's Moon". In: *Icarus* 243 (2014), pp. 173–190.

[28] NASA. *NASA Planetary Data System (PDS) Data Standards: Chapter 6 - Data Set/Data Set Collection Contents and Naming*. Tech. rep. Accessed: 2024-08-12. NASA Planetary Data System (PDS), 2024. URL: `https://https://pds.nasa.gov/datastandards/pds3/standards/sr/Chapter06.pdf`.

[29] NASA. *Diviner Lunar Radiometer Experiment Reduced Data Record and Derived Products Software Interface Specification*. Tech. rep. Accessed: 2024-08-13. NASA Planetary Data System (PDS), 2023. URL: `https://pds-geosciences.wustl.edu/lro/urn-nasa-pds-lro_diviner_derived1/document/Diviner_RDR_SIS.pdf`.

[30] NASA. *PDS Data Product Software Interface Specification (SIS) For Mini-RF Advanced Technologies – Lunar Reconnaissance Orbiter (LRO) Payload Operations Center*. Tech. rep. Accessed: 2024-08-13. NASA, 2010. URL: `https://pds-geosciences.wustl.edu/lro/lro-l-mrflro-4-cdr-v1/lromrf_0001/document/dp_sis/mrflro_dp_sis.pdf`.

[31] NASA. *Moon Mineralogy Mapper DATA PRODUCT SOFTWARE INTERFACE SPECIFICATION*. Tech. rep. Accessed: 2024-08-13. NASA, 2011. URL: `https://pds-imaging.jpl.nasa.gov/documentation/M3_DPSIS.PDF`.

Appendix

A1 Descriptions of included datasets

Diviner [18] - The first dataset presented in this project is from an instrument named the Diviner Lunar Radiometer Experiment (Diviner), currently orbiting the Moon on NASA's Lunar Reconnaissance Orbiter (LRO) which launched on June 18, 2009 [19]. The Diviner instrument is designed to measure surface temperatures on the moon and outputs a single temperature value (in Kelvin) per coordinate.

The GDR files (explained in Appendix A3) contained missing data as a rectangular map was given. This dataset was most affected by interpolation.

For Diviner, files were accessed and processed directly from the web.

LOLA [20] - Similarly to the Diviner dataset, the Lunar Orbiter Laser Altimeter (LOLA) instrument is located on the LRO. The instrument provides several products listed below [21]:

- Global, regional, and local (meterscale) models of geodetic topography, surface slopes, surface roughness, surface brightness
- Improved lunar gravity model

In this project, the measurement of interest is the surface brightness. More specifically it is in units of normal albedo, which is the reflectance of the surface measured at zero phase angle relative to a Lambert surface illuminated at the same geometry [22]. The albedo value of the surface of the moon has been seen to spike at locations where the maxium temperature doesn't exceed the sublimation temperature of water on the moon (110K) [22]. LOLA outputs a single albedo value where 0 is no relfectance and 1 is near complete reflectance, the latter indicating the presence of water ice.

For LOLA, files were accessed and processed directly from the web.

M^3 [23] - The Moon Mineralogy Mapper (M^3) is another NASA instrument however this was flown onboard the Indian Space Research Organisation's (ISRO) Chandrayaan-1 [24]. M^3 was a grating imaging spectrometer that measured spectral reflectance from the from the visible to near-infrared ranges [1]. It was selected to pursue objectives requiring measurement of composition at fine spatial scales over the full lunar surface [25]. Similarly to LOLA, the output from this instrument is spectral data which measures the reflectance of the lunar surface at each point.

This dataset also contained some missing data. This appears to be a region where data was just not gathered, potentially due to the early end of the Chandrayaan-1 mission.

For M^3, files were downloaded first and processed locally. This was necessary as the large number of files led to significant delays when trying to read each one from the web.

Mini-RF dataset [26] - The Miniature Radio Frequency (Mini-RF) is the final dataset considered in this project and is also part of the LRO. The Mini-RF instrument makes synthetic aperture radar observations at two wavelengths and two resolutions [1]. A derived product from the Mini-RF measurements is the circular polarization ratio (CPR), defined as the magnitude of the same sense divided by the opposite sense polarized signals received [27].

This dataset was also downloaded first and processed locally. This was necessary as it compromised a single, extremely large file which could easier be processed locally.

A2 Omitted datasets

A large and thorough analytical study of the resource potential of PSRs was completed by Brown et al. [1]. In that paper, 4 additional datasets to those above were considered when ranking these

regions. Provided below is a brief explanation of these datasets and why they were not included in this project.

LAMP FUV albedo - The Lyman Alpha Mapping Project (LAMP) Far-UltraViolet (FUV) dataset is another dataset measuring reflectance data, this time in the UV range. It was omitted as it contained no north pole data due to low signal to noise ratio (SNR).

Diviner ice depth stability - This is a dataset derived from the Diviner temperature data. It was omitted as the values were based on model predictions rather than direct instrument data. This was deemed unsuitable for use in a deep learning model as trends may be obscured.

LEND and LPNS - The Lunar Exploration Neutron Detector (LEND) and Lunar Prospector Neutron Spectrometer (LPNS) are both instruments which measure data from neutrons to estimate hydrogen abundance on the lunar surface. These instruments had a pixel scale of 10,000 and over 15,000 m/px respectively and this was deemed to poor of a resolution to train a model on as many points would have similar values.

A3   Preprocessing of datasets

The datasets provided by the PDS come as various products, with different levels of processing. A full explanation of the data processing levels is available in NASA's PDS standards [28] with a brief summary provided below.

- PDS3 Level 2: Edited Data - Corrected for telemetry errors and split per instrument. Sometimes refered to as Experimental Data Record, corresponds to NASA Level 0.

- PDS3 Level 4: Resampled Data - Data has been resampled in the time or space domains and could also be calibrated. NASA Level 1B.

- PDS3 Level 5: Derived Data - Derived results, as maps, reports, graphics etc. NASA Levels 2 to 5. Includes Global Data Record (GDR) product.

The highest level available for each dataset was selected due to time constrains and the products used are listed below. Key considerations surrounded whether interpolation had already been applied, as this affects profiling potential inaccuracies in the project where further interpolation is planned.

- Diviner: NASA Level 3 GDR, not interpolated

- LOLA: NASA Level 3 GDR, some interpolation applied

- M^3: NASA Level 2, not interpolated

- Mini-RF: NASA Level 3 GDR, not interpolated.

The 3 LRO datasets used were NASA Level 3 GDR products. The data will have been converted from raw values into scientifically meaningful units, applying calibration coefficients, correcting for known biases and errors in the instrument, correcting for atmospheric effects and projecting onto a specific coordinate system. Unlike the LOLA GDR data products, which use interpolation to create continuous global grids, Diviner GDR data products include data gaps in grid cells where no observations were acquired [29]. The Mini-RF mosaic was generated from multiple Level 2 Calibrated Data Records (CDRs) along with the synthesis of metadata and appears not to have been interpolated as part of the product [30].

For the M^3 dataset, NASA Level 3 products aren't available and so NASA Level 2 Data Products are used. This meant that data was not organised into a global products and had to be assembled manually using location data from L1B Products. As such, there is no risk of points already being interpolated into a regular grid but instead this pre-processing was completed as part of this project [31].

Aside from the additional processing required for the M³ dataset described above, the main processing challenges of all of these datasets was decoding the resource-efficient files on the PDS to obtain a dataframe of the scientific values and accompanying coordinates. This is the bulk of the first section of the project, discussed in Section 4.

A4   Data storage

For the 3 LRO datasets (Diviner, LOLA and Mini-RF), the data was stored in image files (either .img or .jp2 files) with associated label (.lbl) files. The image files contained no headers or other data, all metadata was provided in accompanying label files. All GDRs (LRO datasets) consist of images employing scaled binary 16-byte integers. Image data was converted into a 2D array for processing which was crucial for coordinate calculation. The coordinates of points in the GDRs are according to a specific spatial distribution, dependant on a given projection type.

The M³ datasets were structured slightly differently. Similarly to the GDR products, image files (.IMG) contained no headers or other data and were accompanied by metadata (.LBL) files. M³ images were stored as little endian, 16-bit signed integers, band interleaved by line but these products were, however, converted into 3D arrays as they contained all 86 spectral bands. Coordinates were not dependant on a projection type and were provided in a separate directory where location data was stored as its own image (.IMG) file with accompanying metadata (.lbl).

A5   Dataset details

Table 4: Selected details for the datasets

Dataset	Number of files	Pixel scale (m/px)	Projection Type
Diviner	42	240	Polar Stereographic
LOLA	4	500	Polar Stereographic
M³	887	280	n/a
Mini-RF	1	150 and 30	Equatorial (Simple Cylindrical)

A6   Justification for using Spearman's Rank Correlation and Kendall's Tau to evaluate the effect of the omitted datasets

Spearman's Rank Correlation assesses how well the relationship between the two lists can be described using a monotonic function, making it suitable for understanding overall rank correlation. Kendall's Tau provides a measure of the correspondence between the lists by evaluating the number of concordant and discordant pairs, which offers a more nuanced view of the consistency in ranking order. By applying these metrics, we can quantitatively assess the impact of removing four datasets on the ranking process. If the correlation values are high, it would suggest that the reduced set of factors retains much of the original ranking's integrity, thereby validating the decision to simplify the ranking model without significantly compromising the accuracy or reliability of the results.

A7   Projection conversion equations

For LRO datasets, a projection was provided to turn a standard mesh into coordinates. First, a standard pixel mesh is generated using `np.arange` and calibrated using the sample projection offset, line projection offset and map resolution. Projection types can be seen in Table 4 and Equations 1 and 2 show the conversion into latitude and longitude values.

$$\lambda = \left(\frac{x}{a} \times \frac{360}{2 \times \max(|\lambda|)} + 360 \right) \mod 360 \tag{2a}$$

$$\phi = y \times \frac{\text{max\_lat} - \text{min\_lat}}{\max(|y|)} \tag{2b}$$

Simple Cylindrical Projection Equations

$$\lambda = \left(\text{center\_lon} + \arctan 2(y, x) \times \frac{180}{\pi} \right) \mod 360 \tag{3a}$$

$$\phi = \begin{cases} 90.0° - \frac{c \times 180}{\pi} & \text{if center\_lat} = 90.0° \\ -90.0° + \frac{c \times 180}{\pi} & \text{if center\_lat} = -90.0° \end{cases} \tag{3b}$$

Polar Stereographic Projection Equations

λ : longitude (degrees)
ϕ : latitude (degrees)
x : horizontal pixel coordinate
y : vertical pixel coordinate
a : Moon's radius (km)

c : angular distance from the pole (radians)
center_lon : central longitude of the projection
center_lat : central latitude of the projection
max_lat : maximum latitude
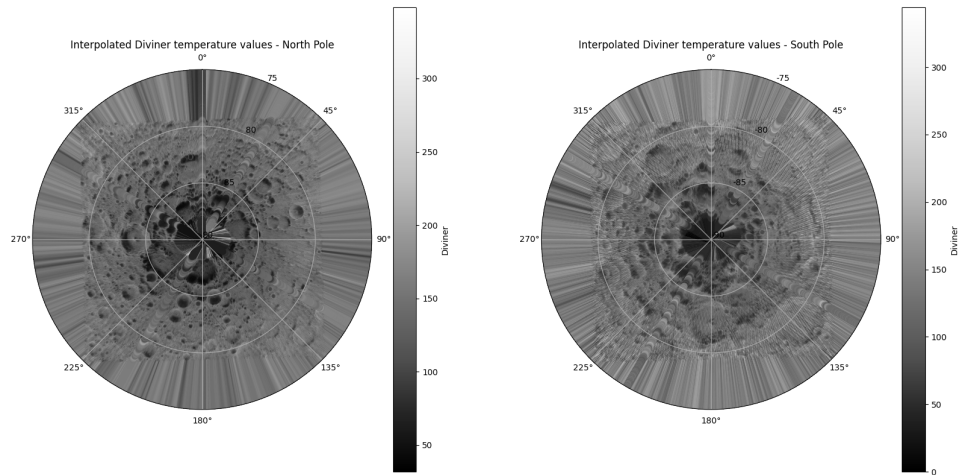min_lat : minimum latitude

A8 Hyperparameter results

Table 5: Hyperparameter tuning results

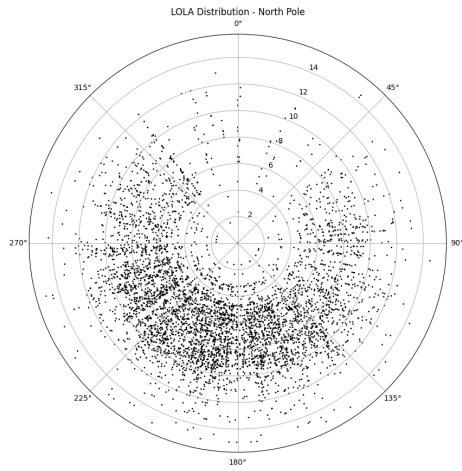| Model	Hidden dimension	Epochs	Learning rate	Dropout rate	Batch size	k	Beta	Weight decay
FCNN	512	2500	2.5e-5	0.3	2^{17}	n/a	0.075	7.5e-4
GCN	512	1000	5e-6	0.75	2^{15}	30	0.25	0.05

A9 Diviner Images
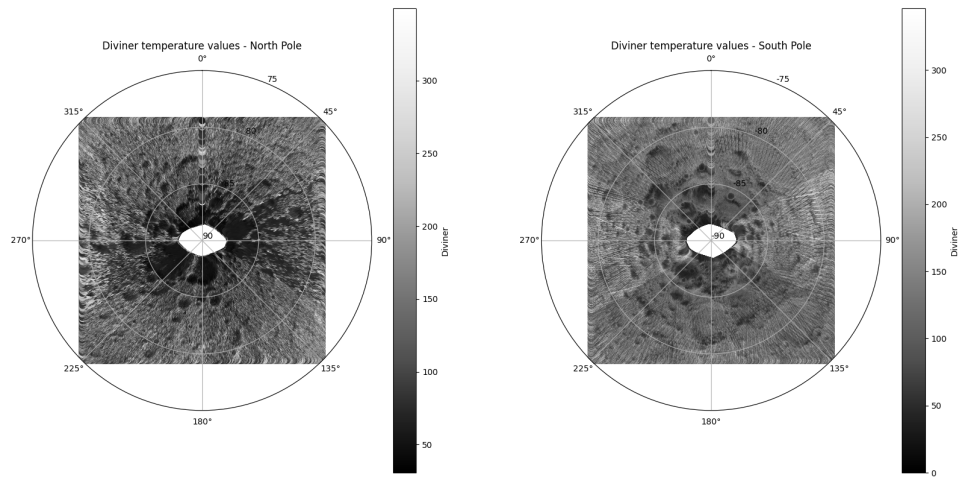
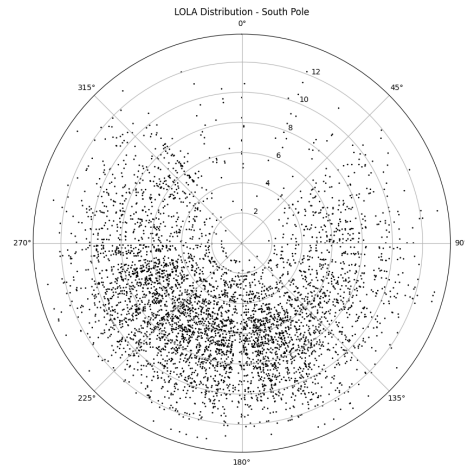(a) Map of raw Diviner values

(b) Map of interpolated Diviner values

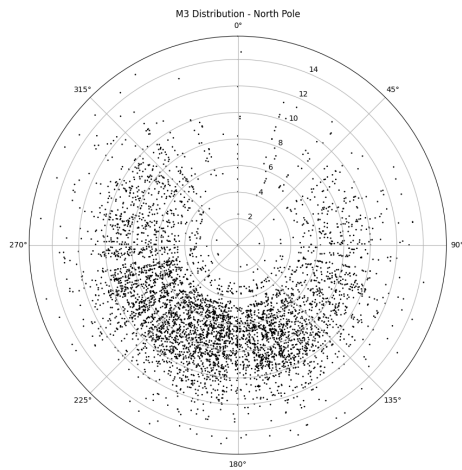Figure 8: The raw Diviner data versus the interpolated data
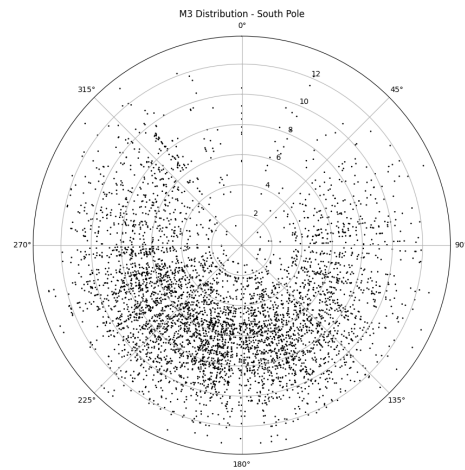
A10 Binary Images

(a) Distribution of values above LOLA threshold
for the North Pole

(b) Distribution of values above LOLA threshold
for the South Pole

(c) Distribution of values above M³ threshold for
the North Pole

(d) Distribution of values above M³ threshold for
the South Pole

Figure 9: Interesting observation regarding the distribution of PSRs. The black dots represent values
above threshold (labelled 1 or 2). Many more points can be seen at longitudes near 180 degrees
compared to 0 degrees. This is likely due to the fact that this side of the Moon constantly faces
away from Earth, making it more susceptible to meteoroid impact. Referenced in Section 4.6.