Comparison between the NDSI threshold detection technique and a Deep Network Model approach: The case of the uncovered glaciers at the Atacama Region, Chile.

Christian Lannefranque
christian.lannefranque@student.uva.nl
Universiteit van Amsterdam
Amsterdam, The Netherlands

Adrien Tavernier
Universidad de Atacama
Copiapo, Chile

Efstratios Gavves
Universiteit van Amsterdam
Amsterdam, The Netherlands

Ayon Garcia
Universidad de Atacama
Copiapo, Chile

Bernard Foing
Universiteit Leiden
Leiden, The Netherlands

Valeria Krzhizhanovskaya
Universiteit van Amsterdam
Amsterdam, The Netherlands

ABSTRACT

Uncovered glacier ice mapping is a current and evolving practice in environmental and cryosphere studies. Snow and ice have a characteristic spectral answer that makes them possible to detect with remote sensing multispectral images and a widely used Normalized Difference Snow Index (NDSI). Here we develop several UNET-based Deep Neural Networks (DNN) facing a segmentation problem to identify and classify uncovered glaciers from remote sensing data, including multispectral data and a digital elevation model. The main issue tackled during this project is related to the fact that remote sensing images have a specific spatial resolution that can produce uncertainty in pixels, especially frequent in the uncovered glacier boundaries, where both ice and non-ice elements are found. To take this uncertainty into account, data sets were splitted in a specific way and extra masks with extended boundaries were produced. Results showed that DNNs performed on par with the baseline: the NDSI Technique. Also, results revealed that leaving certain bands out and extending masks together with elevation data may improve the glacier ice detection capacity. Applying DNN training techniques, blending DNN predictions and trying other multispectral bands input combinations are proposed as future tasks to enhance the DNNs performance for ice segmentation.

KEYWORDS

Glaciers Ice Surface, NDSI, Satellite Remote Sensing, Landsat, Convolutional Deep Neural Network, UNET, Ice Segmentation

1 INTRODUCTION

The Atacama region in northern Chile (see Figure 1), a region with desertic areas in between the Pacific Ocean and the adjacent Andes Mountain range, it is characterized by a semi-arid to hyperarid transition from south to north, where the cryosphere plays a key role in the population's access to fresh water [14]. The glacier ice and permafrost ice reserves located in the mountainous areas contribute, in less quantity but in a more consistent and extensive way compared to snow melt and scarce precipitations, to the water supply of river flows and overall aquifers in the Atacama region [6]. Thereby, economic development and human activity in the Atacama region is tightly related to the behaviour of the region's cryosphere. Considering the actual threats of climate change and global warming, it is imperative to determine the progression of the local cryosphere, whose ice extent and melting rate has a direct consequence in the water available in the valleys.

García et al. (2017) developed an integral inventory of the cryosphere reserves in the Atacama region. The study accounted for a total of 168.908 hectares of ice-reserves, where only 2.7% are represented by uncovered glaciers and the 97.3% corresponds to other cryoforms combining debris-covered glaciers and three permafrost landforms: rock glaciers, gelifluction taluses and protalus lobes. The articles also demonstrated a 34% reduction of the ice inventory from previous measures in 1987 [5]. This study not only serves as a base to understand the cryosphere elements found in the region, their characteristics and history, but also carried out the laborious task of developing manual digital annotations delineating the perimeter and area of all the Cryospheric forms present in the region. This product, as explained later, will be an important asset in the current study since it will be used for the ground truth labelling.

Monitoring of the cryosphere and glaciers in this particular region is a laborious task although of high interest due to its correlation with climate change. Spectral indexes have been developed to facilitate snow detection automatically [17] and, lately, artificial intelligence approaches have also stepped in to perform a more integral and accurate detection of the cryosphere [10].

The objective of this thesis project is to quantify, for the year 2015, the ice surface of the uncovered glaciers linked to the water's contribution of the Atacama Region's watersheds . This estimation is made in two ways. The first and baseline method, widely tested in Remote Sensing Cryosphere Studies, is the Normalized Difference Snow Index (NDSI) threshold method. The second method, based in Artificial Intelligence (AI) and Data Science (DS) studies, is a Deep Neural Network (DNN) approach. The DNN is trained using multispectral remote sensing images, a digital elevation model and Garcia et al.'s manual digitalized annotations of the uncovered ice surface serving as the ground truth. Later, the DNN estimations are analysed and compared to the NDSI threshold technique estimations to drive the final discussion and conclusion points.

1.1 Research Question

This project aims to answer the following research question:

To what extent can a Deep Neural Network approach improve an automated NDSI threshold technique to estimate an uncovered glacier's ice surface?
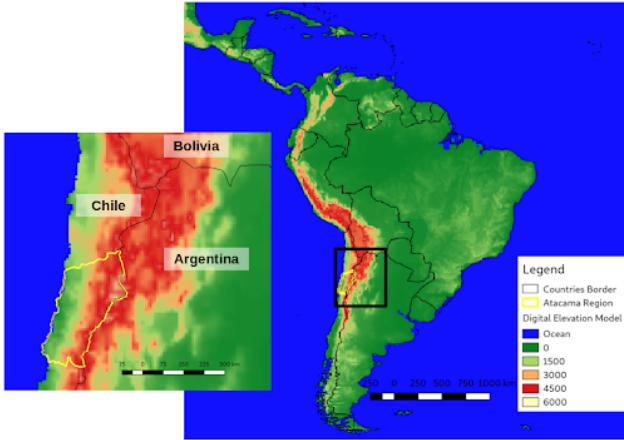
Figure 1: Map and location of the Atacama region.

has been in operation since February 2013 and produces images with a monthly temporal resolution, a 30 meter spatial resolution (900 m^2) and 11 discrete bands associated to different sub ranges of the ES (see figure 2).
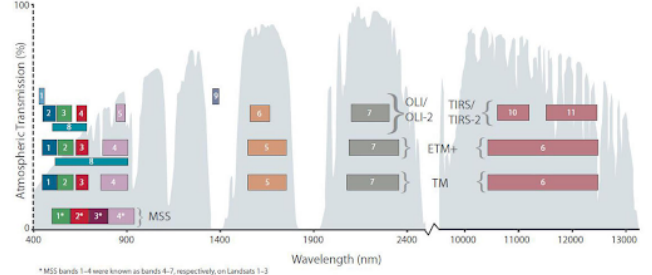

Figure 2: Landsat's Band wavelength allocation [13]

2 RELATED WORK

2.1 Remote Sensing

Remote sensing is the assortment of data about an object or event without having physical contact with the object in question [16]. First remote sensing tools started by the mid 19th century with photographs from air balloons, evolving towards the use of kites, planes and the current frequently used artificial satellites and drones. Remote sensing of the Earth is based on the interaction between the electromagnetic waves and matter. This interaction depends on electromagnetic radiation, i.e. its frequency (or wavelength), and the chemical and physical properties of the observed matter. The range of the electromagnetic radiation frequency, know as the Electromagnetic Spectrum (ES), covers wavelengths from 10-12 m to 1 m which is divided in sub ranges or bands, such as the visible fraction (400 - 700 nm), shortwave infrared (SWIR, 1.4 - 3.0 μm), near infrared (NIR, 700 nm - 1mm), among others. Radiance interacts with matter through absorption, transmission or reflection. Most radiation originates in the sun, reaches the earth, propagates through the atmosphere, interacts with surface matter (soil, vegetation, ocean, etc) and traverses the atmosphere again to finally reach a Satellite Remote Sensor. Thus, satellite remote sensors perceive the radiance or photons (light flux) that interact with the atmosphere after being reflected from the earth surface. Nevertheless, the latest Satellite products apply a post capture atmospheric correction to remove the atmosphere interaction and, therefore, approximate the pure reflectance from the earth surface.

Since the 1970's, space agencies and lately also private organizations, have been developing artificial satellite missions to capture outer space images of the earth. Missions with its different remote sensors, offer products with different features in terms of spatial, temporal, spectral and radiometric resolutions. Spatial resolution indicates the spatial size covered for each image's pixel. Temporal resolution refers to the image's acquisition frequency. Spectral resolution is the capacity to detect different parts of ES. Radiometric resolution refers to discretization of the measured ES. Thus, for example, NASA's Landsat Mission with the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments on board,

2.2 Remote Sensing of the Cryosphere

The cryosphere is the part of the Earth that has water in solid state, being glaciers and snow of particular interest for this study. These different cryosphere elements, although relatively similar to human perception, disclose a different reflectance answer across the ES, known as the spectral signature. Figure 3 shows this spectral behaviour for snow, firn (crystalline or granular snow not yet been compressed into ice) and ice. Thus, the spectral signature can be used to identify these elements using remote sensing images that capture the surface reflectance. Snow has a high reflectance in the visible range and relatively small reflectance in the short wave infrared. Whereas, glaciers are the result of long term snow accumulation with melting/refreezing cycles, thus reflectance response along the ES is reduced due to the presence of impurities and darker elements that compenetrate ice sheets between cycles. Nevertheless, when these cryosphere elements are accompanied by other elements, such as dirt or debris, these can hide and highly distort the expected reflectance [16].

2.2.1 The Normalized Difference Snow Index (NDSI). The Normalized Difference Snow Index (NDSI) is an index used to estimate the presence of snow [9]. As explained before, snow is highly reflective in the visible green range and very absorptive in the NIR and SWIR spectrum. Thus, the NDSI expresses this spectral property as a normalized difference between the visible and SWIR bands (see equation 1).

$$NDSI = \frac{Green - SWIR}{Green + SWIR} \qquad (1)$$

The NDSI can take a value between -1 and 1 and often, with sensitivity study and local domain knowledge, a threshold value is set to indicate the presence of snow. Figure 4 shows an example of the NDSI application.

The studied region is mainly occupied by the Andes mountain soil, which has a very low NDSI value. Proof of this is the distribution of the NDSI values in the studied region (figure 5), where most values are between -0.25 and zero.
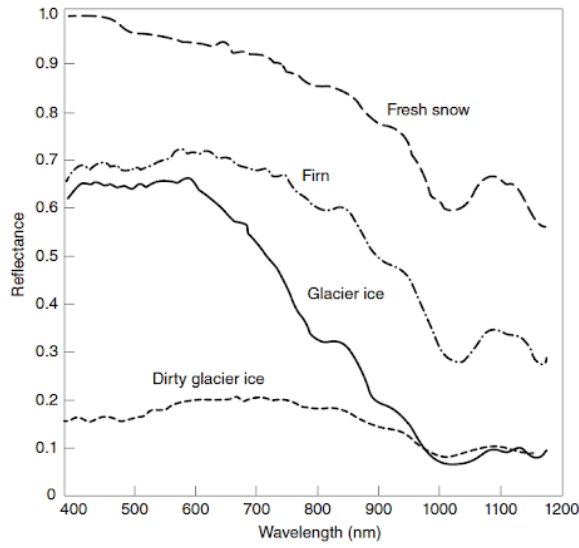
Figure 3: Spectral answer of snow, firn, glacier ice and dirty glacier ice. (Tedesco, 2015).
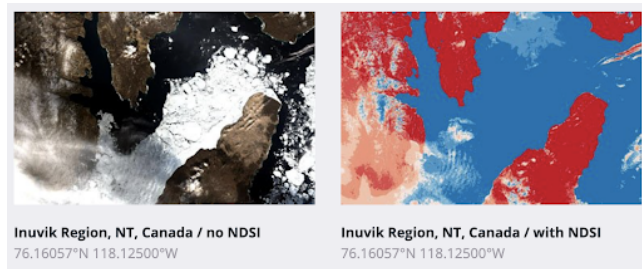

Inuvik Region, NT, Canada / no NDSI
76.16057°N 118.12500°W

Inuvik Region, NT, Canada / with NDSI
76.16057°N 118.12500°W

Figure 4: Example of the NDSI technique applied on a remote sense capture (left) and the resulting image (right) [15]

2.2.2 Related Studies: Glacier mapping in Cordillera Blanca. Measuring a glacier's area has been a task with several years and decades of experience. Cordillera Blanca, a mountain region located in Peru, it's one of the closest glacier regions, in geographical-terms, that partially share topographical characteristics with the Atacama Region, although both are influenced by different climatological patterns. Nevertheless, studies performed in Cordillera Blanca serve as a base to understand current and relevant methodologies for glacier mapping in the Atacama region. Burns Nolin (2013) made estimations of glacier areas, explored the effects of atmospheric elements and topographical corrections in the reflectance measures of debris-free glaciers using satellite multispectral imagery and a NDSI threshold technique. Veettil (2018) also mapped debris-free glaciers in Cordillera Blanca using the NDSI technique to build an inventory of the region. Additionally to the concerning environmental results, both studies revealed important physical and environmental particularities that can influence or mislead the reflectance measures of ice, such as atmospheric particles, the studied area's topography, shadows, clouds, snow and water.
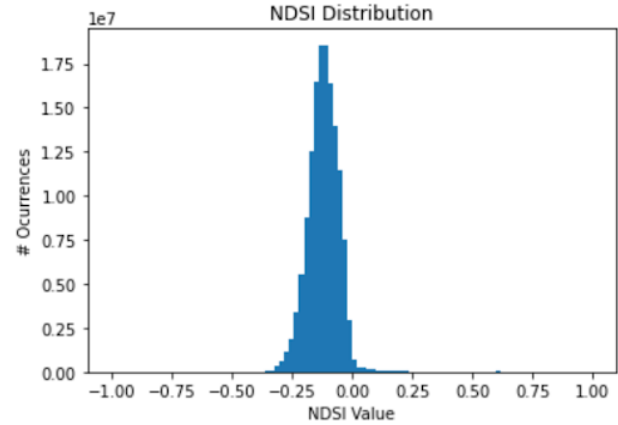

Figure 5: NDSI Value distribution in the Atacama Region.

2.3 Contribution of Artificial Intelligence in Cryosphere mapping

The NDSI technique is commonly used for mapping ice and uncovered glaciers but, since its based on the spectral information of only two bands (Green and SWIR), it may lead to false positive results over elements that partially share the reflectance properties of ice, such as snow, clouds, water and brighter parts of salt flats. Figure 4 shows some of these flaws, as water and clouds are classified as ice. In this way, Artificial Intelligence and specifically Deep Learning approaches, arise as an attractive alternative due to their capabilities of combining several input bands and implementing multiple thresholds through their activation functions.

Authors Yan et al. in their article, propose a glacier discriminator based on a deep learning model [19]. Their model is trained using different spectral bands, reducing noise and having better results compared to other neural networks applied in Sentinel-2 imagery. Another relevant study is the one developed by Robson et al., where a deep learning model, specifically a convolutional neural network (CNN), an object-based image analysis and a digital elevation model (DEM) are used for rock glacier detection [10]. During the study, they recognize that the neural network's capabilities of identifying patterns and shapes can lead to a better segmentation of rock glacier. Their study obtained a 90% detection of the rock glaciers and, although rock glaciers are different from exposed glaciers, the carried out methodology can be applied on other cryospheric landforms and locations.

3 METHODOLOGY

3.1 Data Collection

The first task consisted of gathering the data used for the training, validation and testing of the DNN model. Burns Nolin (2013) and Veettil (2018) articles also provided technical guidelines to consider during this stage of the project. Some of these good practices are (1) a proper selection of satellite scenes to avoid presence of snow and clouds, (2) technical adjustment between satellite resources, (3) understanding of band ratio's techniques, (4) appropriate NDSI threshold selection and (5) approaches for error analysis.

Thus, this particular process involved a reconstruction of the observed region through a careful selection of Landsat 8 multispectral images and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation model, both available from the United States Geological Survey (USGS). Landsat 8 with the OLI and the TIRS sensors, is the last satellite launched in 2013 as part of the Landsat Mission and, compared to prior Landsat instruments, is capable of acquiring deep blue visible and infrared reflectance data [13].

The extraction was done using scenes from 2014 and 2015, since the cryosphere inventory was developed for that same period. The best way to select the specific date, based on Burns Nolin (2013) and Veettil (2018) studies, is to select scenes at the end of the ablation period so as to minimize snow and cloud cover existence. In the Atacama region, this corresponds to the summer period between November (2014) and March (2015). After manual observation in key coordinates, the scenes captured on February 13th, 2015 were selected.

The multispectral images included bands 1 to 7 with a 30 meter pixel resolution and band 10 with a 100 meter pixel resolution, thus encompassing the visible, NIR, SWIR and TIRS spectrum. Adding the ASTER digital elevation model with a 30 meter pixel resolution and one channel with the NDSI calculation, the data considered 10 channels in total.

Finally, the manual digitalized annotations developed by Garcia et al. were used. This resource records the area, perimeter, location, altitude, orientation and other attributes of all the cryospheric landforms of the region: rock glaciers, gelifluction taluses, protalus lobes and uncovered glaciers. The annotations accounted for a total of 356 uncovered glaciers in the Atacama region.

3.2 Dataset Preprocessing

The original scenes included both Chilean and Argentinean territory although the Argentinean fraction had to be removed because the uncovered glacier labeling was only available for the Atacama (Chile) region. Also, few pixel values in the visible spectrum bands were missing due to the low reflectance of dark shaded areas. These missing values were interpolated with valid neighboring pixels using the inverse distance weighting algorithm [4].

At this point, the project encountered its biggest challenge that shaped all upcoming tasks. The main issue faced in this project is related to the pixel resolution and its respective uncertainty. As explained in the related work section, satellite multispectral imagery has a spatial resolution equivalent to the size of a pixel which, for this study, is equivalent to 30 meters. The size of a pixel has therefore an inherent constraint to identify details within an image and, depending on the size of the object in the image (i.e. the amount of pixels where the object is contained), this uncertainty may be relevant or negligible.

In the case of glacier detection, this translates to the fact that many pixels containing a mixture of ice and soil (or other elements) can be found in the data and, its further classification as ice or non-ice pixels, impacts the model's ability to detect the transition zone between the glacier itself and the surrounding soil. This uncertainty is more significant in small glaciers (e.g. glaciers with an area of

about the size of a pixel) and glaciers with an elongated or irregular shape.

Two tasks were established to address the pixel resolution uncertainty. On the one hand, the dataset splits into the training, validation and test sets needed to consider the area and shape of glaciers because, as mentioned before, these two variables are correlated to the glacier's surface uncertainty. Making the splitting in such a way, allows the model's learning process to consider glaciers with different bind uncertainties. On the other hand, to test the uncertainty related to partial ice pixels, models had to be trained with two kinds of masks, one set of masks with the original glacier boundaries and one with a 15 meter outward extension applied to the glacier boundaries. The unmodified boundary mask lets the model consider some blended element pixels as ice and some others as not-ice, whereas the outward extended mask indicates the model to consider all these partial ice pixels as ice. The next paragraphs explain the steps followed in order to comply with these two tasks.

To build the image dataset, the cropping had to be made in such a way that (1) all glaciers have an image where they fit completely and (2) every image is focused on a whole glacier and linked to its corresponding area and shape. *El Potro* glacier, the biggest uncovered glacier of the region with a surface of 6.97 km^2, is contained in 110 pixels width to 115 pixels height. Thus, individual 120 x 120 pixel size images were cropped from the satellite imagery for each of the 356 uncovered glaciers having each glacier in the center of the image. However, with this procedure, images could be overlapped due to glaciers closely located to each other (especially small glaciers). In other words, some glaciers were fully or partially repeated on 2 or more cropped images. This fact was not left unattended because, this meant that glaciers could be repeated across the training, validation and test sets and thus provoke a non-independent test and further biased result [2]. Thus, from the 356 images available, a selection of independent images had to be carried out. The selection process was made prioritizing images with whole glaciers over incomplete or fractional glaciers. This process resulted in a total of 263 glaciers selected across 58 fully independent images.

Correspondingly, masks were built using Garcia et al.'s manual digitized labeling as ground truth to indicate two classes: glacier ice pixels and non glacier ice pixels. As mentioned previously, two sets of masks were built: one set with the original (or also called "unmodified" or "no buffered") ice surface boundaries and one set with a 15 meter outward extension (or also called "buffered mask") of the ice surface boundaries. Additionally, another challenge occurred here: ice pixels are much less frequent than non-ice pixels. Ice pixels account for 6% of all pixels, meaning an unbalanced training scenario for the DNN which in turn, although in general it will produce highly accurate results, it will cause very poor results in terms of glacier ice detection. To address this issue, the DNN will be trained and evaluated with a special Loss function and evaluation metrics. This is explained with more details in the Deep Neural Network section.

Later, the images were splitted into a training, validation and test set, in a 80%, 10% and 10% ratio. Here, to produce a reasonable homogeneous splitting, the distribution of two key descriptive variables linked to the glacier's pixel uncertainty were considered. One variable was the glacier's area and the second variable was

a derived geometric shape factor of the glacier. The latter was calculated as the ratio between the glacier's real perimeter and the perimeter of the circle equivalent to the glacier's surface area. Indeed, the circle is the geometric shape maximising the inner surface with respect to the outer contour and therefore the ideal shape of a glacier in order to limit the uncertainties linked to the pixels located at the soil (outer) - inner (ice) border.

Thus, a glacier with a circular shape will have a geometric shape value close to 1 and a very irregular shaped glacier will tend to have value closer to zero. To understand the underlying distribution of each variable, 93 theoretical distributions were tested using the Kolmogorov-Smirnov test (KS test) to evaluate the goodness of fit. The KS test quantifies the greatest distance between the sample's hypothetical distribution and the theoretical distribution and, under a certain significance level α, rejects or not the null hypothesis regarding the data following a specified distribution [8]. With this, the theoretical distribution with the best fit (p-value) was selected as the underlying distribution for each variable. For the Area, the Power Log Norm distribution was selected and, for the geometric shape factor, the Weibull Max distribution was selected. Figure 6a and 6b show the histogram of both variables and the Probability Mass Function (PMF) of the selected theoretical distribution.

(a)   **(b)**
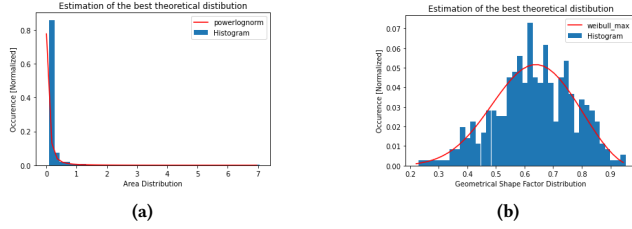
Figure 6: Histogram of Glaciers areas and selected Power Log Norm theoretical distribution PMF (left) and Histogram of Glaciers geometrical shape factor and selected Weibull Max theoretical distribution PMF (right)

To perform the actual splitting a Monte Carlo sensitivity study was pursued. The objective of the Monte Carlo approach is to randomly explore the space of possible sample combinations, find a minimum number of iterations needed to achieve a convergence in the sampling solution and then select the optimal splitting solution discovered. In every iteration, glaciers images are sampled into a preliminary training, validation and test set with a 80%, 10% and 10% proportion. Glacier images are sampled drawing a random value from both the area and the geometrical shape factor theoretical distributions to then select the nearest glacier image, according to the euclidean distance, to the drawn random values. After increasing iterations, the preliminary splitted sets whose distribution fits best to the original distribution (according to a KS test) is selected. Thus, the optimal sampling was found with 900 iterations. This procedure resulted in splitting the 58 independent images into 36 images for training, 12 images for validation and 10 images for testing. The histogram and distribution of both area and geometrical shape factor of each splitted set are shown in table 1.
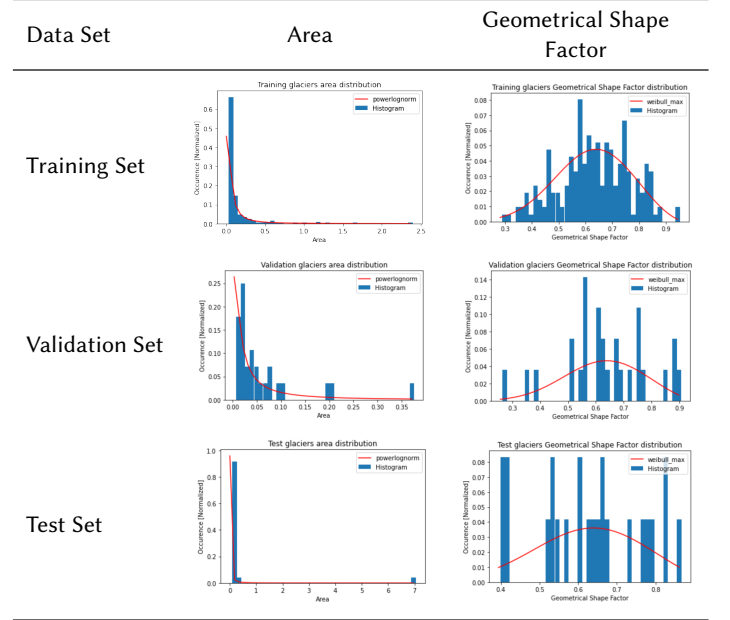

Table 1: Area and Geometrical Shape Factor distribution along Training, Validation and Test sets after glacier selection through Monte Carlo methodology

3.3 Deep Neural Network

To estimate the ice surface, a deep neural network, based on Ronneberger et al.'s pytorch U-NET model, was trained approaching an image segmentation problem. The U-NET architecture was used as in the default configuration. The U-NET architecture (figure 7) consists of 4 blocks containing 2 convolutional layers with batch normalization and ReLU activation function. Max pooling layers are used in the encoding part and up-convolutional layers in the decoding part [3]. Each convolutional layer will transform neighboring data using 3x3 pixel filters [11]. The idea of convolutional layers is to extract local features in an image. The ReLU activation function acts as a non linear transformation after every convolution, taking the maximum between zero and the resultant value of the convolution, thus any resultant negative value is transformed to zero. Max pooling layers reduce the size of the extracted features, by selecting the maximum value within a 2x2 pixel size filter. Max pooling layer's objective is to retain the important features from each region and reduce the image resolution.

The learning rate and epochs were manually selected with the support of empirical experimentation from domain expertise (see table 2).

Table 2: Parameters used for the U-NET Model Training

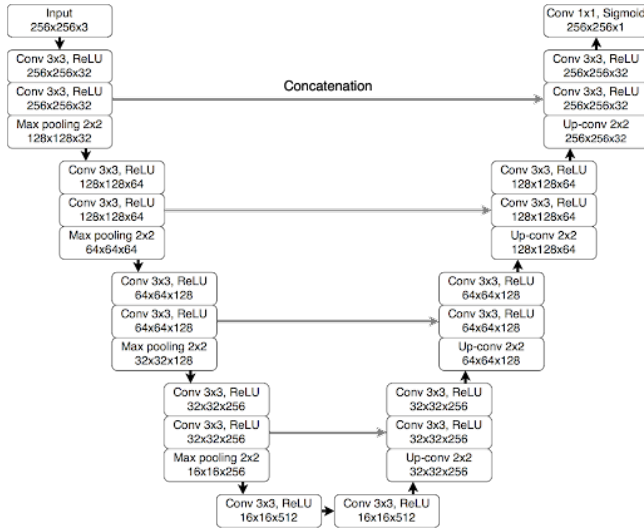| Parameter	Value
Learning Rate	0.01
Epochs	100
Loss Function	Dice Loss

Figure 7: U-NET architecture [3]

For the loss function and, to address the data set's class unbalance problem, the Dice Loss criterion was selected (equation 2) (Reference). The Dice Loss criteria, differently to other widely used criterions like the cross entropy loss, focuses on measuring true positives and thus, conceptually speaking, making the model focus on learning how to identify ice and not how to detect what it is not ice [18]. Input channels were normalized between 0 and 1 using the minimum and maximum value of the respective channel between all the sampled images. Flipping and rotation augmentations were also applied in the images during training stages.

$$DICE = \frac{2|DetectionResult \cap GroundTruth|}{|DetectionResult| + |GroundTruth|} \qquad (2)$$

DNNs were trained under different scenarios in terms of input channel combinations: On the one hand, using the raw inputs, (1) One DNN was trained with all the multispectral bands and the DEM, (2) one with all the multispectral images and (3) one with bands number 3 and number 6. On the other hand, using an assisted NDSI input (based on band 3 and 6), (4) one model was trained only with the NDSI calculation and finally (5) one with the NDSI calculation and the DEM. These same models were then trained using the extended mask, producing a total of 10 models trained. Input combinations 1 and 2 were selected to test all multispectral bands and the DEM. Combinations 3 and 4 were chosen to test whether DNNs are able to decipher additional information from the same bands that the NDSI calculation uses. Combination 5, similar to combination 2, was selected to check if the DEM could add value to the segmentation.

The selection of evaluation metrics considered the context of the unbalanced situation and the need to take into account the costs of both false positives and false negatives. Thus, the following metrics were selected: F1, precision, recall and the total ice surface percentage difference. F1 score measures the model's precision (how much of the predicted ice is really ice) and recall (how much of the real ice is predicted as ice) simultaneously. A model with many

false positives will decrease its precision and a model with many false negatives will decrease its recall. Finally the total ice surface percentage difference is a more broad performance indicator that evaluates the total glacier ice surface prediction.

4 RESULTS

The test set accounted for a total of 10 images with 24 glaciers covering 4.977 km^2 of ice surface. Each DNN model trained, as well as the baseline method: The NDSI Threshold Technique, were then tried in the test set to predict the glacier ice pixels. Table 3 shows the results of each DNN model and the NDSI Threshold Technique

The threshold for the NDSI Threshold Technique was set iterating over different threshold values and selecting the one that maximises the F1 score in the training set. As shown in figure 8, the best F1 score was achieved at a threshold of 0.2 (not generalizable to the rest of the region).
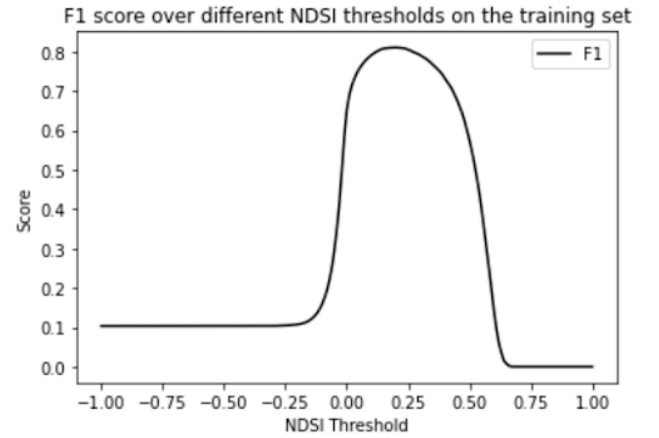

Figure 8: Empirical search of the best threshold to use for the NDSI threshold technique.

5 DISCUSSION

F1 metric will be used as the first metric to analyze when comparing models. Mainly because F1 is a pixel wise metric that, as explained in the methodology section, takes into account both false positives and false negatives. Secondly, considering the context of the Atacama region and the built data set, where most pixels are non-ice pixels, reducing false negatives becomes a priority thus, recall becomes the following metric to evaluate when analyzing a model's performance. The % △ Ice Surface will be the last metric to consider since it's an overall measure that's blind to individual pixel prediction mistakes. Under this definition, model 5 results as the best model using band 6, band 3 and the original mask and achieving an extra 1% over the baseline method.

For further discussion, two particular images were selected from the test set (figure 9a and 9b). Predicted masks over these images were obtained (table 4) and analyzed. The selection criteria was based on the diversity of glaciers that were present in the images and the controversial results produced by the DNN predictions over these. One corresponds to the image of *El Potro* glacier, the biggest

Table 3: Summary table of DNNs trained, the NDSI Threshold Technique and their respective results on the test set

| Model # | Input Type | Input Channels | Mask Type | Statistical Metrics | | | | Physical Metrics | |
				F1	Balanced Accuracy	Precision	Recall	Predicted Ice Surface [km^2]	% △ Ice Surface
1	Raw	All Bands + DEM	Original	0.83	0.90	0.86	0.81	4.648	-6.60%
2		All Bands	Original	0.87	0.92	0.90	0.85	4.682	-5.93%
5		B3 + B6	Original	0.88	0.92	0.91	0.85	4.660	-6.37%
6		All Bands + DEM	15 M Extention	0.86	0.93	0.85	0.87	5.109	2.66%
7		All Bands	15 M Extention	0.79	0.92	0.75	0.85	6.188	24.32%
10		B3 + B6	15 M Extention	0.84	0.93	0.81	0.86	5.750	15.53%
3	Assisted	NDSI	Original	0.87	0.95	0.85	0.90	5.306	6.62%
4		NDSI + DEM	Original	0.86	0.94	0.84	0.88	5.206	4.61%
8		NDSI	15 M Extention	0.82	0.92	0.80	0.85	5.813	16.80%
9		NDSI + DEM	15 M Extention	0.84	0.90	0.87	0.81	5.072	1.92%
11	Raw	Baseline: NDSI Threshold Technique		0.87	0.94	0.86	0.88	5.110	2.68%

glacier in the data set. The second image corresponds to an area where many kinds of glaciers and other elements could be found: small glaciers, medium size glaciers, elongated glaciers and even snow.

(a) Selected Test Image 1 (El Potro glacier)

(b) Selected Test Image 2

Figure 9: Selected test images. Black pixels correspond to Argentinean territory (removed).

In the upcoming discussion points, please also refer to table 4 to follow the analysis made.

5.1 DNN NDSI Models (Models 3 & 8) vs DNN B3 + B6 Models (Models 5 & 10)

DNN models that interact with both bands 3 and 6 independently have a slight advantage, in terms of the F1 metric, over the ones that use the assisted NDSI input which, as a reminder, is a pixel wise calculation equal to the normalized difference between band 3 and 6. However, when using the original mask recall metric drops and precision increases whereas when training with the extended mask both recall and precision increases. Test images reflect this behaviour: On the one hand, DNNs working with the original mask and the NDSI input tend to have a less restrictive classification, thus producing more false positives in the borders of *El Potro* glacier and in isolated snow patches of the second test image but, at the same time, identifying small glaciers in the same image. Conversely, DNNs working with the original mask and bands 3 and 6 independently seem to have a more "conservative" prediction and therefore avoiding false positives in *El Potro*'s border and snow patches in the second test image but losing sight of the small glacier in the same area. On the other hand, both DNNs working with the extended

mask produce false positives at *El Potro*'s border and in small snow patches but also being able to identify the small glaciers in the second test image.

In the case of the original mask, this behaviour may be explained due to the fact that the NDSI input is a pixel wise (or intra-pixel) pre calculation given to the DNN (value closer to 1 for snow or ice and closer to -1 otherwise). This implies that DNNs using this input (1) have a smaller space to try different linear combinations between the underlying bands and (2) lose the capacity to consider the underlying band's data of neighboring pixels in inner layer neurons and during convolutional steps. Therefore, DNNs based in the NDSI may be induced to train based on an intra-pixel pre given information whereas DNNs using the raw bands have the chance to produce a more balanced linear combination considering both intra-pixel data and inter-pixel data. In partial-ice pixel locations such as in *El Potro*'s border and in small glaciers, neighboring pixel's data may be crucial to classify a pixel as ice or not. Thus probably, DNNs based in the raw bands, classify *El Potro*'s border and small glaciers as non-ice due to high presence of surrounding soil and the Argentinean territory set to zero.

In the case of the extended mask, results may be explained due to a high influence of the extension of borders and thus training models to classify partial-ice pixels as ice. This doesn't oppose to the previous hypothesis, since extending mask borders is done regardless of the neighboring pixels and only promotes a less restrictive classification of ice in uncertain pixels.

5.2 Digital Elevation Model

The digital elevation model (DEM), in the scenarios trained with the original mask, DNN's performance worsened. However, using the 15 meter extended masks, the DEM made a positive impact on most of the performance indicators. This might be explained due to elevation contrasts, like cliffs or drops between ice and soil, present in some lower bound glacier boundaries. Therefore, models using the 15 meter extended mask may be able to incorporate this altitude variation factor to discern ice from soil and detect boundaries.

The DEM original purpose was to be used as a discriminator for hard negative elements. Uncovered glaciers are only present over 4.000 meters above sea level and many of them are located in sloping areas. Lakes and bright parts of salt flats, with a similar spectral response to ice and snow, are present at lower altitudes and/or have a slight slope or none. Thus the DEM input was expected to used

to detect altitude gradients and altitude thresholds to discern un-covered glaciers from hard negative elements. Unfortunately, these kinds of samples were not finally present in the dataset, therefore this hypothesis could not be tested.

5.3 Original mask vs extended mask

Extended masks deteriorated F1 results for the majority of the models, mainly due to a drop in their precision, i.e. producing many false positives. This, as mentioned in previous points, is explained due to extended masks being more permissive in classifying partial-ice pixels as ice, thus promoting the models to predict ice pixels in areas where a mixture of ice and debris is present. Neverthe-less, model 1 and 6, trained with all multispectral bands and the digital elevation model, benefited from the extended mask. This model was able to fairly maintain its precision, boost its recall by 6 percentage points and reach the other model's F1 performance with the original masks. Selected test images show a big impact in the case of *El Potro* glacier where even some inner glacier pixels, far from the glacier's border, are wrongly classified as non-ice but corrected when extended masks are added in the equation. This isolated case may be explained due to the slope of *El Potro* glacier. The area of *El Potro* glacier closer to the Argentinean territory (black pixels in the images) is characterized for having a big cliff and pronounced slopes. This altitude gradient, as explained in the previous discussion point, may be a crucial variable for DNNs to detect ice-soil borders which, in this particular case, can influence DNN 1 to classify this sloping area as soil. Then, this wrong classi-fication is solved by the introduction of the extended mask, thus suggesting that the extended mask is a more relevant factor than the DEM input although both bring valuable information for ice segmentation.

5.4 DNN Models vs NDSI Threshold Technique

In general terms, DNN models were able to perform as well as the baseline method: The NDSI Threshold Technique. Particularly, model 5 increased the F1 score by a slight 1% and model 6 performed on par including the % △ Ice Surface physical metric. Model 5, a model with a more prudent decision making as seen earlier, has a 5% improvement in the precision metric and decrease of 3% in the recall metric compared to the NDSI threshold technique. Selected test images show that the NDSI threshold technique is less restrictive in areas where partial-ice pixels are more frequent: glacier borders, small glaciers and elongated glaciers, thus producing false positives of surrounding snow for example, whereas model 5 has a more prudent decision making and classifies these controversial pixels as non-ice. However, the NDSI threshold technique is able to detect small glaciers where model 5 is not even able to detect a slight portion of them due to its same conservative behaviour.

This difference in results may be explained by the pixel-wise nature of the NDSI of Threshold Technique compared to the multi-pixel and non-linear combinations performed by DNNs 5 and 6. Furthermore, the NDSI Threshold Technique can be assimilated to a single neuron, where the neuron's inputs are both band 3 and band 6 values for a single pixel, the normalized difference of bands 3 and 6 is the calculation done in the neuron and the threshold corresponds to a ReLU activation function. Thus it is easier to see that the

NDSI Threshold Technique lacks neighboring pixels information whereas DNNs 5 and 6 can include this data into the non-linear combinations. This means that DNNs conservative classification is influenced by the neighboring data surrounding the pixel being classified. Therefore, these DNNs are prone to classify small glaciers or bordering pixels in bigger glaciers as non-ice because of the non-ice surrounding pixels, whereas the pixel by pixel decision making of the NDSI threshold technique is not influenced by neighboring pixels.

DNN model 6 produced very similar results in the total ice sur-face estimation: just 1 km^2 of difference with the NDSI Threshold Technique. But this metric is tricky, since although the overall result is very close to the baseline method, it is opaque when looking at more detailed results in the selected test images: The DNN models still produce false positives in *El Potro* glacier's border and also false negatives related to small glaciers or elongated glaciers that aren't identified in the second test image. Thus, when looking at the overall sum of the glacier ice surface, these two mistakes are compensated.

In this case, the false positive results of DNN 6 may be explained by the use of the extended mask (explained in section 5.3) and the false negative results may be due to the high frequency of non-ice pixels around very small glaciers.

5.5 Input Analysis

An input analysis was carried out to better understand the chan-nel's relevance in the DNN models. For this, one DNN model was selected to predict the glacier ice surface over a singular test image several times but, in every inference, one of the channels is hidden on purpose (set to zero) to check how results change. The model selected for this experiment was model 6, since it's the model with the best F1 score using all multispectral bands and the DEM. The specific test image used for this is the second test image (figure 9b). The selection was based on the controversial false positive and false negative results that it previously produced and the fact that it includes relatively big glaciers, small glaciers, and elongated shaped glaciers. Table 5 shows the results of this experiment.

The results of the input analysis reveal that band 1 is the most relevant band, especially to avoid false negatives that are easy to fix and thus to achieve a decent recall score. Band 6 and 10 are crucial bands to avoid false positives and thus maintaining the model's precision. Interestingly, hiding band 7 may improve all evaluation metrics of ice detection, finding hard glacier shapes like small glaciers, elongated glaciers or even uncovered glaciers with a low reflective response.

In order to obtain more complete spectral information and to return to a physical basis, we used data from the Hyperion demon-strator instrument on board the EO-1 probe [12] which delivers hyperspectral imagery. In hyperspectral mode, the data collected allow us to benefit from a much finer spectral resolution than in multispectral mode (242 bands for Hyperion compared to 8 for OLI in the VIS-NIR-SWIR range) and thus to get closer to theoretical spectra.

As Hyperion was a demonstrator, the spatial coverage only cov-ers a few areas of the globe and only includes level 1 data, i.e.
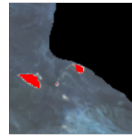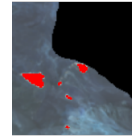
Test Image	Ground Truth	NDSI Threshold Technique	DNN 1	DNN 3	DNN 5	DNN 6	DNN 8	DNN 10
1 (El Potro)								
2								

Table 4: Ground Truth mask and predicted masks by the NDSI Threshold Technique and the DNN models in the selected test images

top atmosphere. In order to benefit from surface reflectance, atmospheric effects were corrected using a code derived from the QUAC algorithm [1], using experimental soil spectra present in the Cordillera [7].

A Hyperion acquisition made in the region of interest dating from March 2004 was chosen, allowing the selection of a spectral response of the ice from two glaciers in the Atacama region and a spectral response characteristic of the soils in the Cordillera (figure 10).
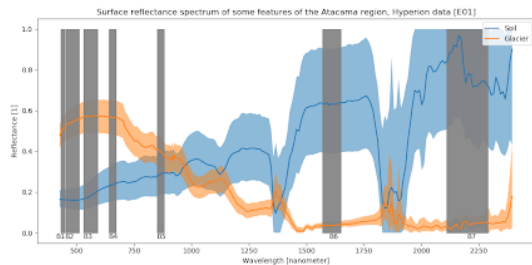
Figure 10: Sample hyperspectral answer of glaciers and soil in the Atacama region.

The results of band 1 being selected as the most relevant band may be due to the relatively similar spectral answer of both ice and soil between band 1 and band 3, considering band 3 as one the decisive bands to detect ice. Band 1 may also be an important factor not only to identify ice, but to identify what is not ice, but this should be tested in further tasks. To understand the influence of band 10 and 7 further DNNs testing should be done an analyzed.

5.6 Limitations

The main limitation in the project is the traceability and explainability of the decision making done by the Deep Neural Network models. It is imperative, especially in physics and environmental studies, to be able to interpret the DNN internal processes. Nowadays, many analyses are being developed and tested in order to

understand DNNs better but these kinds of studies were postponed in order to focus on the pixel resolution issue.

A second limitation, although as important as the first one, is related to the data set construction. The pixel resolution problematic implicated a glacier-based cropping, which left out other natural elements (salt flats, lakes, snow, etc) present in the region that, had they been included in the data set, would have produced a more robust and generalizable DNN model. Also, this cropping approach led to a low amount of images available in the data set (58), thus limiting the DNNs learning capacity during training.

6 CONCLUSION

To conclude, it was possible to found Deep Neural Networks that could perform as well as to the baseline method: the traditional NDSI Threshold Technique. Model 5 which, using the same spectral bands as in the NDSI Threshold method, increased the F1 score in 1% by improving its precision. Model 6 using all the multispectral bands, a digital elevation model and extended masks, showed similar performance.

Analysis over the results may indicate that pre calculated inputs (such as the NDSI) may impact the DNNs ability to extract further inter-pixel spectral information. In terms of mask types, results showed that some models might take advantage of these extended masks and the DEM information to discriminate fuzzy borders between ice and soil and thus find a way to bypass the uncertainty of partial ice pixels. Nevertheless, extended mask should be used carefully: they may increase small glacier detection but also increase the false positive rate. The input analysis indicated the existence of more relevant bands and even maybe bands producing noise. Thus, an adequate selection emphasizing certain bands and withdrawing others, may lead to better results.

As a proposition, DNNs could be used in a more strategical way. Considering the societal and vital importance of the relative downstream water flux produced by glaciers, the cost and impact of making a false positive classification in the borders of a big glacier, such as *El Potro*, is much bigger than making a false positive classification of small and/or isolated glaciers. Thus it could be beneficial to use a more conservative DNN model, such a DNN 5, to
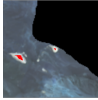
Hidden Channel	Predicted Mask	F1 Score	Precision	Recall
None		0.72	0.98	0.57
B1		0.25	1	0.14
B2		0.55	1	0.38
B3		0.56	1	0.38
B4		0.70	0.98	0.60
B5		0.71	0.99	0.56
B6		0.45	0.30	0.95
B7		0.8	0.75	0.87
B10		0.34	0.20	0.98
DEM		0.57	1	0.40

Table 5: Inputs analysis results

specifically discern a big glacier's border and use a less restrictive DDN model, such as DNN 10, to detect small glaciers.

The pixel spatial resolution uncertainty issue was the biggest challenge that this project tried to embrace. This challenge required big emphasis in finding a methodology to develop an homogeneous data set splitting, so models could be trained, validated and tested under the best possible circumstances so DNNs could learn from all kinds of glacier shapes and sizes and thus, their linked pixel uncertainty. This procedure, although it generated certain disadvantages, reduced this uncertainty to a minimum which, in a contrary simplistic random or uniform data set split approach, may have led to poorer and unreliable results.

To answer the research question: To what extent can a Deep Neural Network approach improve an automated NDSI threshold technique to estimate an uncovered glacier's ice surface? It is possible to indicate that Deep Neural Networks can match the performance of the NDSI Threshold Technique to estimate an uncovered glacier's ice surface and will likely improve with further works. It's undeniable that the NDSI Threshold Technique already achieves very reasonable results within the context of the simple problem of ice detection whose spectral answer has been very well studied. Nevertheless it is possible to expect that DNNs will be able to produce results in a more dynamic and generalizable way, thus being adaptable to different locations and other periglacial landforms.

6.1 Future Work

One next step is to include sample images with hard negative elements (salt flats and water). A second suggested step would be to train DNNs with input bands where performance was benefited, such as with band 1, 3, 6, the DEM and the extended mask. A third and relatively straight forward task would be to hyperparameter tune the DNN models and set an early stop methodology in the training stages. Last but not least, combining DNN predictions to produce a final result may improve the glacier ice surface segmentation.

ACKNOWLEDGMENTS

REFERENCES

[1] Larry Bernstein, Steven Adler-Golden, Xuemin Jin, Brian Gregor, and Robert Sundberg. 2012. Quick atmospheric correction (QUAC) code for VNIR-SWIR spectral imagery: Algorithm details. *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, 1–4. https://doi.org/10.1109/WHISPERS. 2012.6874311
[2] Andrew Brown. 2020. Fundamentals of Data Science.
[3] Mateusz Buda, Ashirbani Saha, and Maciej Mazurowski. 2019. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine* 109 (2019). https://doi.org/10.1016/j.compbiomed.2019.05.002
[4] Geospatial Foundation. [n.d.]. GDAL Grid Tutorial. https://gdal.org/tutorials/ gdal_grid_tut.html
[5] Ayon Garcia, Christopher Ulloa, Gonzalo Amigo, and Juan PAblo Milana. 2017. An inventory of cryospheric landforms in the arid diagonal of South America (high Central Andes, Atacama region, Chile). *Quaternary International* 438 (may 2017), 4–19. https://doi.org/10.1016/j.quaint.2017.04.033
[6] Christian Hunter. 2015. A dynamic, multivariate sustainability measure for robust analysis of water management under climate and demand uncertainty in an arid environment. *Water* 7, 11 (2015), 5928–5958.
[7] Clark R.N. Swayze G.A. Livo K.E. Hoefen T.M. Pearson N.C. Wise R.A. Benzel W.M. Lowers H.A. Driscoll R.L. Kokaly, R.F. and A.J. Klein. 2017. USGS Spectral

Library Version 7: U.S. Geological Survey Data Series. *IEEE Transactions on Geoscience and Remote Sensing* 1035, 61 (2017). https://doi.org/10.3133/ds1035.

[8] National Institute of Standards and Technology. [n.d.]. Kolmogorov-Smirnov Goodness-of-Fit Test. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

[9] George Riggs and Dorothy K. Hall. 2015. *MODIS Snow Products Collection 6 User Guide.* NSIDC.

[10] Benjamin Robson, Tobias Bolch, Shelley MacDonell, Daniel Holbling, Philipp Rastnere, and Nicole Schaffer. 2020. Automated detection of rock glaciers using deep learning and object-based image analysis. *Remote Sensing of Environment* 250 (December 2020). https://doi.org/10.1016/j.rse.2020.112033

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV]

[12] J. A. Mendenhall S. G. Ungar, J. S. Pearlman and D. Reuter. 2003. Overview of the earth observing one (EO-1) mission. *IEEE Transactions on Geoscience and Remote Sensing* 41, 6 (June 2003), 1149–1159. https://doi.org/10.1109/TGRS.2003.815999

[13] United States Geological Service. [n.d.]. USGS EROS Archive - Landsat Archives - Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor)

Level-1 Data Products. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-operational-land-imager-and

[14] Francisco Suarez. 2014. Integrated water resource management and energy requirements for water supply in the Copiapó river basin, Chile. *Water* 6, 9 (2014), 2590–2613.

[15] Earth Observing System. [n.d.]. NDSI. https://eos.com/ndsi/

[16] M. Tedesco. 2015. *Remote Sensing of the Cryosphere.* Wiley Blackwell.

[17] Bijeesh Veettil. 2018. Glacier mapping in the Cordillera Blanca, Peru, tropical Andes, using Sentinel-2 and Landsat data. *Singapore Journal of Tropical Geography* (2018). https://doi.org/10.1111

[18] Qinyang Lu Xiaoling Xia and Xin Gu. 2019. Exploring An Easy Way for Imbalanced Data Sets in Semantic Image Segmentation. *Journal of Physics* 1213 (2019). https://doi.org/10.1088/1742-6596/1213/2/022003

[19] Shuai Yan, Linlin Xu, and Rui Wu. 2019. Automatic Classification of Glaciers from Sentinel-2 Imagery Using A Novel Deep Learning Model. *ICAIP 2019: Proceedings of the 2019 3rd International Conference on Advances in Image Processing* (November 2019), 155–159. https://doi.org/10.1111